Rate Region of the Vector Gaussian One-Helper Source-Coding Problem

Md. Saifur Rahman and Aaron B. Wagner*

Abstract

We determine the rate region of the vector Gaussian one-helper source-coding problem under a covariance matrix distortion constraint. The rate region is achieved by a simple scheme that separates the lossy vector quantization from the lossless spatial compression. The converse is established by extending and combining three analysis techniques that have been employed in the past to obtain partial results for the problem.

Keywords: multiterminal source coding, one-helper problem, covariance matrix distortion constraint, vector Gaussian sources, vector quantization, distortion projection, source enhancement.

1 Introduction

We study the vector Gaussian one-helper source-coding problem[1], depicted in Fig. 1. Here $\mathbf{X}$ and $\mathbf{Y}$ are two jointly vector Gaussian sources. Encoders 1 and 2 observe two i.i.d. strings distributed according to $\mathbf{X}$ and $\mathbf{Y}$, respectively, and separately send messages to the decoder at rates R_1 and R_2 bits per observation, respectively, using noiseless channels. The decoder uses both messages to estimate $\mathbf{X}$ such that a given distortion constraint on the average error covariance matrix is satisfied. The goal is to determine the rate region of the problem, which is the set of all rate pairs (R_1, R_2) that allow us to satisfy the distortion constraint for some design of the encoders and the decoder.

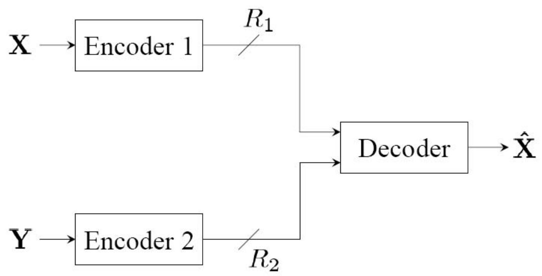

Figure 1: Vector Gaussian one-helper source-coding problem.

Oohama [1] gave a complete characterization of the rate region for the case in which both sources are scalar. His achievability proof is a Gaussian scheme that is described in more detail below. The converse argument uses the entropy-maximizing property of the Gaussian distribution and the entropy power inequality (EPI), and it bears a certain resemblance to Bergmans' earlier converse for the scalar Gaussian broadcast channel [2]. As such, one might hope that the *channel enhancement* technique introduced by Weingarten *et al.* [3] to solve the MIMO Gaussian broadcast channel would be sufficient to solve the problem considered here. This turns out not to be the case, however. Among other contributions, Liu and Viswanath [4] showed that *channel enhancement* yields an outer bound for the vector one-helper problem that is not tight in general. This was later improved slightly by the present authors to show that the Gaussian scheme achieves a portion of the boundary of the rate region [5]. Liu and Viswanath's approach was later subsumed by Zhang [6], who applied enhacement in a different way and called it *source enhancement*, but this also yielded an outer bound that is not always tight.

*Both authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. (Email: mr534@cornell.edu, wagner@ece.cornell.edu.)

[1]The material in this paper was presented in part at the 49th Annual Allerton Conference on Communications, Control, and Computing, University of Illinois, Urbana-Champaign, Sept. 2011.

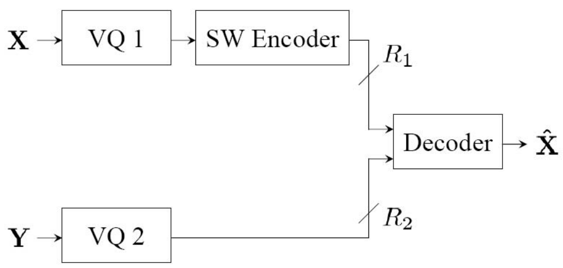

Figure 2: A Gaussian achievable scheme.

The case in which Y is a scalar and $\mathbf{X}$ is a vector was recently solved by the authors [7]. The proof did not use enhancement, but it did require a novel technique that we call *distortion projection.* Here we shall show that *distortion projection*, *source enhancement*, and Oohama's converse technique together are sufficient to solve the general problem in which both $\mathbf{X}$ and $\mathbf{Y}$ are vectors. In particular, we shall determine the rate region exactly and show that a vector extension of the Gaussian scheme used by Oohama is optimal. In this scheme, as depicted in Fig. 2, encoder 1 vector quantizes (VQ) its observations using a Gaussian test channel as in point-to-point rate-distortion theory. It then compresses the quantized values using Slepian-Wolf (SW) encoding [8]. Encoder 2 just vector quantizes its observations using another Gaussian test channel. The decoder decodes the quantized values and estimates the observations of encoder 1 using a minimum mean-squared error (MMSE) estimator.

The rest of the paper is organized as follows. Section 2 explains the notation used in the paper. In Section 3, we present the mathematical formulation of the problem, a description of the scheme, and the statement of our main result. Section 4 gives an outline of the converse argument. Since the proof of the converse is somewhat involved, it is divided into Sections 5 through 8.

2 Notation

We use uppercase to denote random variables and vectors. Boldface is used to distinguish vectors from scalars. Arbitrary realizations of random variables and vectors are denoted in lowercase. For a random vector $\mathbf{X}$, $\mathbf{X}^n$ denotes an i.i.d. vector of length n, $\mathbf{X}^n(i)$ denotes its *i*th component, and $\mathbf{X}^n(i:j)$ denotes the *i*th through *j*th components. The superscript T denotes matrix transpose. The covariance matrix of $\mathbf{X}$ is denoted by $\mathbf{K_X}$. The conditional covariance matrix of $\mathbf{X}$ given $\mathbf{Y}$ is denoted by $\mathbf{K_{X|Y}}$ and is defined as

$$\mathbf{K_{X|Y}} \triangleq E\left[(\mathbf{X} - E(\mathbf{X}|\mathbf{Y}))\,(\mathbf{X} - E(\mathbf{X}|\mathbf{Y}))^T\right].$$

All vectors are column vectors and are m-dimensional, unless otherwise stated. We use $\mathbf{I}_m$ to denote an $m \times m$ identity matrix. With a little abuse of notation, $\mathbf{0}$ is used to denote both zero vectors and zero matrices of appropriate dimensions. We use $\text{Diag}(d_1, d_2, \ldots, d_p)$ to denote a diagonal matrix with diagonal entries $d_1, d_2, \ldots, d_p$. The trace of a matrix $\mathbf{A}$ is denoted by $\text{Tr}(\mathbf{A})$. For two real symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \succcurlyeq \mathbf{B}$ ($\mathbf{A} \succ \mathbf{B}$) means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite (definite). Similarly, $\mathbf{A} \preccurlyeq \mathbf{B}$ ($\mathbf{A} \prec \mathbf{B}$) means that $\mathbf{B} - \mathbf{A}$ is positive semidefinite (definite). All logarithms in this paper are to the base 2. The determinant of a matrix $\mathbf{K}$ is denoted by $|\mathbf{K}|$. The notation $X \leftrightarrow Y \leftrightarrow Z$ means that X, Y, and Z form a Markov chain in this order. We use $\text{span}\{\mathbf{c}_i\}_{i=1}^l$ to denote the subspace spanned by $\{\mathbf{c}_i\}_{i=1}^l$.

3 Problem Formulation and Main Results

Let $\mathbf{X}$ and $\mathbf{Y}$ be two generic zero-mean jointly Gaussian random vectors with covariance matrices $\mathbf{K_X}$ and $\mathbf{K_Y}$, respectively. Initially, we shall assume that $\mathbf{X}$ is m-dimensional and $\mathbf{Y}$ is k-dimensional. Let $\{(\mathbf{X}^n(i), \mathbf{Y}^n(i))\}_{i=1}^n$ be a sequence of i.i.d. random vectors with the distribution at a single stage being the same as that of the generic pair $(\mathbf{X}, \mathbf{Y})$. As depicted in Fig. 1, encoder 1 observes $\mathbf{X}^n$ and sends a message to the decoder using an encoding function

$$f_1^{(n)} : \mathbb{R}^{mn} \mapsto \left\{1, \ldots, M_1^{(n)}\right\}.$$

Analogously, encoder 2 observes $\mathbf{Y}^n$ and sends a message to the decoder using another encoding function

$$f_2^{(n)} : \mathbb{R}^{kn} \mapsto \left\{1, \ldots, M_2^{(n)}\right\}.$$

The decoder uses both received messages to estimate $\mathbf{X}^n$ using a decoding function

$$g^{(n)} : \left\{1, \ldots, M_1^{(n)}\right\} \times \left\{1, \ldots, M_2^{(n)}\right\} \mapsto \mathbb{R}^{mn}.$$

Definition 1. *A rate-distortion vector* $(R_1, R_2, \mathbf{D})$ *is* achievable *for the vector Gaussian one-helper source-coding problem if there exist a block length* n*, encoding functions* $f_1^{(n)}$ *and* $f_2^{(n)}$*, and a decoding function* $g^{(n)}$ *such that*

$$R_i \geq \frac{1}{n} \log M_i^{(n)} \textit{ for all } i \in \{1, 2\}, \textit{ and}$$
$$\mathbf{D} \succcurlyeq \frac{1}{n} \sum_{i=1}^{n} E\left[\left(\mathbf{X}^n(i) - \hat{\mathbf{X}}^n(i)\right)\left(\mathbf{X}^n(i) - \hat{\mathbf{X}}^n(i)\right)^T\right],$$

where

$$\hat{\mathbf{X}}^n \triangleq g^{(n)}\left(f_1^{(n)}\left(\mathbf{X}^n\right), f_2^{(n)}\left(\mathbf{Y}^n\right)\right).$$

Let $\mathcal{RD}$ *be the set of all achievable rate-distortion vectors and* $\overline{\mathcal{RD}}$ *be its closure. Define*

$$\mathcal{R}(\mathbf{D}) \triangleq \left\{(R_1, R_2) : (R_1, R_2, \mathbf{D}) \in \overline{\mathcal{RD}}\right\}.$$

We call $\mathcal{R}(\mathbf{D})$ *the* rate region *for the vector Gaussian one-helper source-coding problem.*

Our goal is to characterize the rate region $\mathcal{R}(\mathbf{D})$. Note that the matrix distortion constraint is more general in the sense that it subsumes other natural distortion constraints such as a finite number of upper bounds on the mean square error of reproductions of linear functions of the source. In particular, it subsumes the case in which the distortion constraint is on the mean square error of reproductions of the components of $\mathbf{X}$.

Since we are interested in a quadratic distortion constraint, without loss of generality we can restrict the decoding function to be the MMSE estimate of $\mathbf{X}^n$ based on the received messages. Therefore, $\hat{\mathbf{X}}^n$ can be written as

$$\hat{\mathbf{X}}^n = E\left[\mathbf{X}^n \middle| f_1^{(n)}\left(\mathbf{X}^n\right), f_2^{(n)}\left(\mathbf{Y}^n\right)\right].$$

We can assume without loss of generality[2] that

$$\mathbf{X} = \mathbf{Y} + \mathbf{N},$$

where $\mathbf{N}$ is a zero-mean Gaussian random vector with the covariance matrix $\mathbf{K_N}$ and is independent of $\mathbf{Y}$. The case in which $\mathbf{K_X} \preccurlyeq \mathbf{D}$ has a trivial solution. In this case, the rate region is the entire nonnegative quadrant. So, we assume that $\mathbf{K_X} \preccurlyeq \mathbf{D}$ does not hold in the rest of the paper. This means that there exists a direction $\mathbf{z} \neq \mathbf{0}$ such that

$$\mathbf{z}^T \mathbf{K_X} \mathbf{z} > \mathbf{z}^T \mathbf{D} \mathbf{z}. \tag{1}$$

For now, we assume that $\mathbf{K_X}, \mathbf{K_Y}$, and $\mathbf{D}$ are positive definite. The general case of the problem will be addressed in Section 8.

[2]Since $\mathbf{X}$ and $\mathbf{Y}$ are jointly Gaussian, we can write

$$\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{N},$$

where $\mathbf{A}$ is an $m \times k$ matrix and $\mathbf{N}$ is an m-dimensional zero-mean Gaussian random vector that is independent of $\mathbf{Y}$. Since there is no distortion constraint on $\mathbf{Y}$, and $\mathbf{AY}$ is a sufficient statistic for $\mathbf{X}$ given $\mathbf{Y}$ (i.e., $\mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{AY}$ and $\mathbf{X} \leftrightarrow \mathbf{AY} \leftrightarrow \mathbf{Y}$), we can relabel $\mathbf{AY}$ as $\mathbf{Y}$ and write

$$\mathbf{X} = \mathbf{Y} + \mathbf{N}.$$

3.1 Rate Region

The rate region $\mathcal{R}(\mathbf{D})$ is a closed convex set in the nonnegative quadrant. It is closed by definition and is convex because any convex combination of two points in the rate region is in the rate region as it can be achieved by time-sharing between the encoding and decoding strategies of the two points. Therefore, we can characterize it completely by its supporting hyperplanes, which can be expressed as the following optimization problem

$$\mathcal{R}(\mathbf{D}, \mu) \triangleq \inf_{(R_1, R_2) \in \mathcal{R}(\mathbf{D})} \mu R_1 + R_2,$$

where μ is a nonnegative real number. Let us define

$$\mathcal{R}^*(\mathbf{D}, \mu) \triangleq \begin{cases} v\left(P_{pt-pt}\right) & \text{if } 0 \leq \mu \leq 1 \\ v\left(P_{G1}\right) & \text{if } \mu > 1, \end{cases}$$

where $v\left(P_{pt-pt}\right)$ and $v\left(P_{G1}\right)$ are the optimal values of the optimization problems (P_{pt-pt}) and (P_{G1}), respectively, which are defined as

$$\begin{aligned} (P_{pt-pt}) \quad \triangleq \quad & \min_{\mathbf{K}_{\mathbf{X}|\mathbf{U}}} \quad \frac{\mu}{2} \log \frac{|\mathbf{K}_{\mathbf{X}}|}{|\mathbf{K}_{\mathbf{X}|\mathbf{U}}|} \\ & \text{subject to} \quad \mathbf{K}_{\mathbf{X}} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U}} \succcurlyeq \mathbf{0} \quad \text{and} \\ & \qquad\qquad\quad \mathbf{D} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U}}, \end{aligned}$$

and

$$\begin{aligned} (P_{G1}) \quad \triangleq \quad & \min_{\mathbf{K}_{\mathbf{Y}|\mathbf{V}}, \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}}} \quad \frac{\mu}{2} \log \frac{|\mathbf{K}_{\mathbf{Y}|\mathbf{V}} + \mathbf{K}_{\mathbf{N}}|}{|\mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}}|} + \frac{1}{2} \log \frac{|\mathbf{K}_{\mathbf{Y}}|}{|\mathbf{K}_{\mathbf{Y}|\mathbf{V}}|} \\ & \text{subject to} \quad \mathbf{K}_{\mathbf{Y}} \succcurlyeq \mathbf{K}_{\mathbf{Y}|\mathbf{V}} \succcurlyeq \mathbf{0}, \\ & \qquad\qquad\quad \mathbf{K}_{\mathbf{Y}|\mathbf{V}} + \mathbf{K}_{\mathbf{N}} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} \succcurlyeq \mathbf{0}, \quad \text{and} \\ & \qquad\qquad\quad \mathbf{D} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}}. \end{aligned}$$

We use similar notation to denote other optimization problems and their optimal values throughout the paper. The main result of this paper is the following theorem.

Theorem 1. *The minimum weighted sum rate for the vector Gaussian one-helper source coding problem is given by the solution to the above matrix optimization problem*

$$\mathcal{R}(\mathbf{D}, \mu) = \mathcal{R}^*(\mathbf{D}, \mu).$$

3.2 A Gaussian Achievable Scheme

In this subsection, we present a Gaussian achievable scheme (Fig. 2). The scheme is well-known and is sometimes referred to as the Berger-Tung scheme [9, 10]. This scheme is known to be optimal for several problems in Gaussian multiterminal source-coding literature [1, 11, 12, 13, 14, 15, 16]. However, it is not optimal in some cases. For instance, a lattice-based scheme can outperform it if the goal is to reconstruct a hidden random vector that is jointly Gaussian with $\mathbf{X}$ and $\mathbf{Y}$ [17, 18], and the discrete memoryless version of the scheme can be suboptimal if the sources have common components [19]. For the problem under consideration however, we shall prove that the Berger-Tung scheme is indeed optimal. We present an overview of the scheme here. The details for similar problem setups can be found in [1, 11].

Let $\mathcal{S}$ be the set of zero-mean jointly Gaussian random vectors $\mathbf{U}$ and $\mathbf{V}$ such that

(C1) $\mathbf{U}, \mathbf{X}, \mathbf{Y}$, and $\mathbf{V}$ form a Markov chain $\mathbf{U} \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{V}$, and

(C2) $\mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} \preccurlyeq \mathbf{D}$.

Consider any $(\mathbf{U}, \mathbf{V}) \in \mathcal{S}$ and a large block length n. Let $R_1^{'} \triangleq I(\mathbf{X}; \mathbf{U}) + \epsilon$, where $\epsilon > 0$. To construct the codebook for encoder 1, first generate $2^{nR_1^{'}}$ independent codewords $\mathbf{U}^n$ randomly according to the marginal distribution of $\mathbf{U}$, and then uniformly distribute them into 2^{nR_1} bins. Encoder 2's codebook is constructed by generating 2^{nR_2} independent codewords $\mathbf{V}^n$ randomly according to the marginal distribution of $\mathbf{V}$.

Given a source sequence $\mathbf{X}^n$, encoder 1 looks for a codeword $\mathbf{U}^n$ that is jointly typical with $\mathbf{X}^n$, and sends the index b of the bin to which $\mathbf{U}^n$ belongs. Encoder 2, upon observing $\mathbf{Y}^n$, sends the index of the codeword $\mathbf{V}^n$ that is jointly typical with $\mathbf{Y}^n$. The decoder receives the two indices, then looks into the bin b for a codeword $\mathbf{U}^n$ that is jointly typical with $\mathbf{V}^n$. The decoder can recover $\mathbf{U}^n$ and $\mathbf{V}^n$ with high probability as long as

$$\begin{aligned} R_1 &\geq I(\mathbf{X};\mathbf{U}|\mathbf{V}) \quad \text{and} \\ R_2 &\geq I(\mathbf{Y};\mathbf{V}). \end{aligned}$$

The decoder then computes the MMSE estimate of the source $\mathbf{X}^n$ given the messages $\mathbf{U}^n$ and $\mathbf{V}^n$, and (C2) above guarantees that this estimate will satisfy the covariance matrix distortion constraint. Let

$$\begin{aligned} \mathcal{R}_G(\mathbf{D}) \triangleq \big\{(R_1, R_2) : \ & \text{there exists } (\mathbf{U},\mathbf{V}) \in \mathcal{S} \text{ such that} \\ & R_1 \geq I(\mathbf{X};\mathbf{U}|\mathbf{V}) \quad \text{and} \\ & R_2 \geq I(\mathbf{Y};\mathbf{V})\big\}. \end{aligned}$$

Furthermore, define

$$\mathcal{R}_G(\mathbf{D},\mu) \triangleq \min_{(R_1,R_2)\in\mathcal{R}_G(\mathbf{D})} \mu R_1 + R_2.$$

The following lemma gives the weighted sum-rate achieved by this scheme.

Lemma 1. *The Gaussian achievable scheme achieves $\mathcal{R}_G(\mathbf{D},\mu)$ and*

$$\mathcal{R}_G(\mathbf{D},\mu) = \mathcal{R}^*(\mathbf{D},\mu).$$

Proof. It follows immediately that the Gaussian achievable scheme achieves $\mathcal{R}_G(\mathbf{D},\mu)$. The equality in Lemma 1 is proved in Appendix A. □

Lemma 1 implies that

$$\mathcal{R}(\mathbf{D},\mu) \leq \mathcal{R}^*\left(\mathbf{D},\mu\right).$$

We prove the reverse inequality (converse) next. Since the proof is rather long, we divide it into sections. The next section gives a nonrigorous overview of the argument. In the following section, we study the optimization problem (P_{G1}) in the definition of $\mathcal{R}^*(\mathbf{D},\mu)$ and establish several properties that its optimal solution satisfies. We use these properties in Section 6 to prove the main result needed for the converse. We finally complete the proof of Theorem 1 in Section 7.

4 Overview of the Converse Argument

The starting point of our proof is Oohama's converse for the scalar case, which proceeds as follows. Let $f_1^{(n)}$ and $f_2^{(n)}$ be encoding functions and $g^{(n)}$ be a decoding function that achieve the rate-distortion vector (R_1, R_2, D). Let $C_1 \triangleq f_1^{(n)}(X^n)$ and $C_2 \triangleq f_2^{(n)}(Y^n)$. By standard steps, we have

$$\begin{aligned} nR_2 &\geq \log M_2^{(n)} \\ &\geq H(C_2) \\ &= I(Y^n; C_2). \end{aligned}$$

Likewise, we have

$$\begin{aligned} nR_1 &\geq \log M_1^{(n)} \\ &\geq H(C_1) \\ &\geq H(C_1|C_2) \\ &= I(X^n; C_1|C_2) \\ &= I(X^n; C_1, C_2) - I(X^n; C_2). \end{aligned}$$

It follows that

$$\begin{aligned} nR_1 \geq \quad & \inf_{C_1,C_2} \quad I(X^n;C_1,C_2) - I(X^n;C_2) \\ & \text{subject to} \quad \sum_{i=1}^{n} E\left[(X^n(i) - E[X^n(i)|C_1,C_2])^2\right] \leq nD, \\ & \qquad\qquad I(Y^n;C_2) \leq nR_2, \quad \text{and} \\ & \qquad\qquad X^n \leftrightarrow Y^n \leftrightarrow C_2. \end{aligned} \tag{2}$$

Now this infimum can be lower bounded by separately optimizing each term

$$nR_1 \geq \quad \begin{aligned} & \inf_{C_1,C_2} \quad I(X^n;C_1,C_2) \\ & \text{subject to} \quad \sum_{i=1}^{n} E\left[(X^n(i) - E[X^n(i)|C_1,C_2])^2\right] \leq nD \end{aligned} \quad - \quad \begin{aligned} & \sup_{C_2} \quad I(X^n;C_2) \\ & \text{subject to} \quad I(Y^n;C_2) \leq nR_2 \quad \text{and} \\ & \qquad\qquad X^n \leftrightarrow Y^n \leftrightarrow C_2. \end{aligned} \tag{3}$$

The first optimization problem,

$$\begin{aligned} & \inf_{C_1,C_2} \quad I(X^n;C_1,C_2) \\ & \text{subject to} \quad \sum_{i=1}^{n} E\left[(X^n(i) - E[X^n(i)|C_1,C_2])^2\right] \leq nD, \end{aligned}$$

which we call the *distortion problem*, can be solved using the entropy-maximizing property of the Gaussian distribution and the concavity of the logarithm. The second problem,

$$\begin{aligned} & \sup_{C_2} \quad I(X^n;C_2) \\ & \text{subject to} \quad I(Y^n;C_2) \leq nR_2 \quad \text{and} \\ & \qquad\qquad X^n \leftrightarrow Y^n \leftrightarrow C_2, \end{aligned} \tag{4}$$

which we call the *helper problem*, can be solved via the conditional version of the entropy power inequality [2]. Substituting these solutions into (3) yields exactly the R_1 achieved by the scheme from the previous section for the given R_2 and D. This completes Oohama's converse proof for the scalar case.

The key to Oohama's proof is that separately minimizing the two terms in (2) does not decrease the objective. More precisely, for any pair (C_1^*, C_2^*) that achieves the infimum in (2) we have

$$\begin{aligned} I(X^n;C_1^*,C_2^*) = \quad & \inf_{C_1,C_2} \quad I(X^n;C_1,C_2) \\ & \text{subject to} \quad \sum_{i=1}^{n} E\left[(X^n(i) - E[X^n(i)|C_1,C_2])^2\right] \leq nD, \end{aligned} \tag{5}$$

and

$$\begin{aligned} I(X^n;C_2^*) = \quad & \sup_{C_2} \quad I(X^n;C_2) \\ & \text{subject to} \quad I(Y^n;C_2) \leq nR_2 \quad \text{and} \\ & \qquad\qquad X^n \leftrightarrow Y^n \leftrightarrow C_2, \end{aligned} \tag{6}$$

Whenever (5) occurs, we shall say that the *distortion problem incurs no loss.* Whenever (6) occurs, we shall say that the *helper problem incurs no loss.*

It is not difficult to verify that this proof also works when X is a scalar and $\mathbf{Y}$ is a vector. In particular, both the distortion and helper problems incur no loss in this case. When both $\mathbf{X}$ and $\mathbf{Y}$ are vectors, the proof breaks down in three places:

1. The distortion problem incurs a loss in general. For instance, if $\mathbf{D} \preccurlyeq \mathbf{K_X}$, then the distortion problem is solved by choosing C_1 and C_2 so that

$$\sum_{i=1}^{n} E\left[\Big(\mathbf{X}^n(i) - E[\mathbf{X}^n(i)|C_1,C_2]\Big)\Big(\mathbf{X}^n(i) - E[\mathbf{X}^n(i)|C_1,C_2]\Big)^T\right] = n\mathbf{D}.$$

That is, the constraint is met with equality. For the original problem in (2), on the other hand, even if $\mathbf{D} \preccurlyeq \mathbf{K_X}$ we can only guarantee that

$$\sum_{i=1}^{n} E\left[\left(\mathbf{X}^n(i) - E[\mathbf{X}^n(i)|C_1^*, C_2^*]\right)\left(\mathbf{X}^n(i) - E[\mathbf{X}^n(i)|C_1^*, C_2^*]\right)^T\right] \preccurlyeq n\mathbf{D},$$

and equality does not hold in general. The lack of equality is easiest to see when $\mathbf{K_Y}$ is poorly conditioned. If $\mathbf{K_Y}$ has essentially one nonzero eigenvalue, then the helper will allocate all of its rate in the direction of the associated eigenvector. If R_2 is large, this could result in "overshooting" the distortion constraint in that direction.

2. The helper problem also incurs a loss in general. One way of seeing this is to note that if the goal is only to maximize the mutual information in (4), then one might choose C_2 to favor a direction along which the distortion constraint $\mathbf{D}$ is not active over one for which it is. This would necessarily deviate from the optimizer C_2^* of the original problem.

3. The vector EPI does not solve the helper problem in general.

To address the first issue, observe that the distortion problem incurs no loss if the optimizers C_1^* and C_2^* for the original problem happen to meet the distortion constraint with equality, i.e., it holds that

$$\sum_{i=1}^{n} E\left[\left(\mathbf{X}^n(i) - E[\mathbf{X}^n(i)|C_1^*, C_2^*]\right)\left(\mathbf{X}^n(i) - E[\mathbf{X}^n(i)|C_1^*, C_2^*]\right)^T\right] = n\mathbf{D}.$$

In prior work [7], we showed that it is possible to reduce the general case to this one by projecting the source and the distortion constraint in the directions in which the distortion constraint is met with equality for the candidate optimal scheme. We call this process *distortion projection.* This addresses the first issue. One can verify that if $\mathbf{X}$ is a vector and Y is a scalar, then the second and third issues do not arise, and hence *distortion projection* together with Oohama's converse arguments is sufficient to solve the problem [7].

Liu and Viswanath [4] showed that the *channel enhancement* technique of Weingarten *et al.* [3] is sufficient to solve the helper problem in the vector case, thereby addressing the third issue. Their solution, however, is not sufficient to handle the second issue. Recently, Zhang [6] introduced a variation on the enhancement idea called *source enhancement* that subsumes Liu and Viswanath's approach. *Source enhancement* effectively replaces the original problem with a relaxation for which the helper problem incurs no loss and the vector EPI solves the helper problem, although Zhang does not describe it in this way. This addresses the second and third issues. Thus it appears that *distortion projection*, *source enhancement*, and Oohama's converse technique together should be sufficient to solve the case in which both $\mathbf{X}$ and $\mathbf{Y}$ are vectors. We shall show that this is indeed true. *Source enhancement* and Oohama's converse technique are lifted directly from [1, 6]. The *distortion projection*, on the other hand, requires an extension beyond what was needed in the scalar helper case [7]. This extension requires us to first establish several properties of the optimal Gaussian solution to the problem, to which we turn next.

5 Properties of the Optimal Gaussian Solution

In this section, we study the optimization problem (P_{G1}) defined in Section 3.1. Note first that the constraints

$$\mathbf{K_{Y|V}} \succcurlyeq \mathbf{0} \quad \text{and}$$
$$\mathbf{K_{X|U,V}} \succcurlyeq \mathbf{0}$$

are never active because otherwise the objective value is infinite. We therefore ignore these constraints in the study of the problem. Now, instead of studying (P_{G1}) directly as it is, we study an equivalent formulation. This formulation is implicit in [6]. Note that if $\mathbf{K_{Y|V}}$ and $\mathbf{K_{X|U,V}}$ are feasible for (P_{G1}), then there exist two positive semidefinite matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ such that

$$\begin{aligned}
\mathbf{K_{Y|V}} &= \mathbf{K_Y} - \mathbf{B}_2, \\
\mathbf{K_{X|U,V}} &= \mathbf{K_{Y|V}} + \mathbf{K_N} - \mathbf{B}_1 \\
&= \mathbf{K_Y} - \mathbf{B}_2 + \mathbf{K_N} - \mathbf{B}_1 \\
&= \mathbf{K_X} - \mathbf{B}_1 - \mathbf{B}_2, \quad \text{and}
\end{aligned}$$
$$\mathbf{K_X} - \mathbf{B}_1 - \mathbf{B}_2 \preccurlyeq \mathbf{D}.$$

Therefore, (P_{G1}) is equivalent to the following problem

$$(P_{G2}) \quad \triangleq \quad \min_{\mathbf{B}_1,\mathbf{B}_2} \quad \frac{\mu}{2}\log\frac{|\mathbf{K_X}-\mathbf{B}_2|}{|\mathbf{K_X}-\mathbf{B_1}-\mathbf{B}_2|}+\frac{1}{2}\log\frac{|\mathbf{K_Y}|}{|\mathbf{K_Y}-\mathbf{B}_2|}$$
$$\text{subject to} \quad \mathbf{B}_i \succcurlyeq \mathbf{0} \ \text{ for all } \ i\in\{1,2\}, \ \text{ and}$$
$$\mathbf{D} \succcurlyeq \mathbf{K_X}-\mathbf{B_1}-\mathbf{B_2}.$$

We next establish several properties that the optimal solution to (P_{G2}) satisfies.

Since (P_{G2}) has continuous objective and a compact feasible set, there exists an optimal solution $(\mathbf{B}_1^*,\mathbf{B}_2^*)$ to it. The Lagrangian of the problem is [20, Sec. 5.9.1]

$$\frac{\mu}{2}\log\frac{|\mathbf{K_X}-\mathbf{B}_2|}{|\mathbf{K_X}-\mathbf{B_1}-\mathbf{B}_2|}+\frac{1}{2}\log\frac{|\mathbf{K_Y}|}{|\mathbf{K_Y}-\mathbf{B}_2|}-\mathrm{Tr}\Big(\mathbf{B}_1\mathbf{M}_1+\mathbf{B}_2\mathbf{M}_2-(\mathbf{K_X}-\mathbf{B}_1-\mathbf{B}_2-\mathbf{D})\mathbf{\Lambda}\Big),$$

where $\mathbf{M}_1,\mathbf{M}_2$, and $\mathbf{\Lambda}$ are positive semidefinite Lagrange multiplier matrices corresponding to the constraints $\mathbf{B}_1\succcurlyeq\mathbf{0},\mathbf{B}_2\succcurlyeq\mathbf{0}$, and $\mathbf{D}\succcurlyeq\mathbf{K_X}-\mathbf{B_1}-\mathbf{B_2}$, respectively. The KKT conditions for this problem are [20, Sec. 5.9.2]

$$\frac{\mu}{2}(\mathbf{K_X}-\mathbf{B}_1^*-\mathbf{B}_2^*)^{-1}-\mathbf{\Lambda}^*-\mathbf{M}_1^*=\mathbf{0}, \tag{7}$$

$$\frac{\mu}{2}(\mathbf{K_X}-\mathbf{B}_1^*-\mathbf{B}_2^*)^{-1}-\frac{\mu}{2}(\mathbf{K_X}-\mathbf{B}_2^*)^{-1}+\frac{1}{2}(\mathbf{K_Y}-\mathbf{B}_2^*)^{-1}-\mathbf{\Lambda}^*-\mathbf{M}_2^*=\mathbf{0}, \tag{8}$$

$$\mathbf{B}_i^*\mathbf{M}_i^*=\mathbf{0}, \quad \text{for all} \quad i\in\{1,2\} \tag{9}$$

$$(\mathbf{K_X}-\mathbf{B}_1^*-\mathbf{B}_2^*-\mathbf{D})\mathbf{\Lambda}^*=\mathbf{0}, \quad \text{and} \tag{10}$$

$$\mathbf{M}_1^*,\mathbf{M}_2^*,\mathbf{\Lambda}^*\succcurlyeq\mathbf{0}, \tag{11}$$

where $\mathbf{M}_1^*,\mathbf{M}_2^*$, and $\mathbf{\Lambda}^*$ are optimal Lagrange multiplier matrices. Conditions (7) and (8) respectively are obtained by setting gradients of the objective with respect to $\mathbf{B}_1$ and $\mathbf{B}_2$ to zero. Conditions (9) through (10) are slackness conditions on the Lagrange multiplier matrices. We next establish that these KKT conditions must hold at $(\mathbf{B}_1^*,\mathbf{B}_2^*)$.

Lemma 2. *There exist matrices* $\mathbf{M}_1^*,\mathbf{M}_2^*$, *and* $\mathbf{\Lambda}^*$ *that satisfy the KKT conditions (7) – (11).*

Proof. See Appendix B. □

Let us define

$$\mathbf{\Delta}^*\triangleq\mathbf{\Lambda}^*-\frac{\mu}{2}\left[(\mathbf{K_X}-\mathbf{B}_1^*-\mathbf{B}_2^*)^{-1}-(\mathbf{K_X}-\mathbf{B}_2^*)^{-1}\right].$$

It follows from conditions (7) and (8) that

$$\mathbf{\Delta}^*=\frac{\mu}{2}(\mathbf{K_X}-\mathbf{B}_2^*)^{-1}-\mathbf{M}_1^*=\frac{1}{2}(\mathbf{K_Y}-\mathbf{B}_2^*)^{-1}-\mathbf{M}_2^*. \tag{12}$$

We have the following lemma.

Lemma 3. $\mathbf{\Delta}^*$ *is a nonzero positive semidefinite matrix.*

Proof. See Appendix C. □

If $\mathbf{\Delta}^*$ happens to be positive definite, then *distortion projection* turns out to be unnecessary. To handle the case in which $\mathbf{\Delta}^*$ is singular, we shall use *distortion projection*. Since $\mathbf{\Delta}^*,\mathbf{M}_1^*$, and $\mathbf{M}_2^*$ are positive semidefinite, we can write their spectral decompositions as

$$\mathbf{\Delta}^*=\sum_{i=1}^{r}\lambda_i\mathbf{s}_i\mathbf{s}_i^T, \tag{13}$$

$$\mathbf{M}_1^*=\sum_{i=1}^{p}\alpha_i\mathbf{a}_i\mathbf{a}_i^T, \quad \text{and} \tag{14}$$

$$\mathbf{M}_2^*=\sum_{i=1}^{q}\beta_i\mathbf{b}_i\mathbf{b}_i^T, \tag{15}$$

where

(i) $0 < r \leq m$,

(ii) $0 \leq p, q \leq m$,

(iii) $\lambda_i > 0$, for all $i \in \{1, \ldots, r\}$,

(iv) $\alpha_i > 0$, for all $i \in \{1, \ldots, p\}$

(v) $\beta_i > 0$, for all $i \in \{1, \ldots, q\}$, and

(vi) $\{\mathbf{s}_i\}_{i=1}^r, \{\mathbf{a}_i\}_{i=1}^p$, and $\{\mathbf{b}_i\}_{i=1}^q$ are sets of orthonormal vectors.

Note that we allow p and q to be zero because $\mathbf{M}_1^*$ and $\mathbf{M}_2^*$ can be zero. Since (12) implies

$$\begin{aligned}\boldsymbol{\Delta}^* + \mathbf{M}_1^* &= \frac{\mu}{2}(\mathbf{K_X} - \mathbf{B}_2^*)^{-1} \succ \mathbf{0} \quad \text{and} \\ \boldsymbol{\Delta}^* + \mathbf{M}_2^* &= \frac{1}{2}(\mathbf{K_Y} - \mathbf{B}_2^*)^{-1} \succ \mathbf{0},\end{aligned}$$

we must have

$$\begin{aligned}r + p &\geq m \quad \text{and} \\ r + q &\geq m.\end{aligned}$$

This means that if $r + p = m$, then $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_r, \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p$ must be linearly independent. Similarly, if $r + q = m$, then $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_r, \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_q$ must be linearly independent.

Define the matrix

$$\mathbf{S} \triangleq \left[\sqrt{\lambda_1}\mathbf{s}_1, \sqrt{\lambda_2}\mathbf{s}_2, \ldots, \sqrt{\lambda_r}\mathbf{s}_r\right].$$

It now follows from the definition of $\boldsymbol{\Delta}^*$ that

$$\boldsymbol{\Lambda}^* \succcurlyeq \boldsymbol{\Delta}^* = \mathbf{S}\mathbf{S}^T$$

because

$$(\mathbf{K_X} - \mathbf{B}_1^* - \mathbf{B}_2^*)^{-1} \succcurlyeq (\mathbf{K_X} - \mathbf{B}_2^*)^{-1}.$$

This and (10) imply that

$$(\mathbf{K_X} - \mathbf{B}_1^* - \mathbf{B}_2^* - \mathbf{D})\mathbf{S} = \mathbf{0}. \tag{16}$$

Let $\mathbf{C}$ be an $m \times m$ positive definite matrix and $\{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_t\}$ be a set of $m \times m$ positive definite matrices.

Definition 2. *A non-zero $m \times p$ matrix $\mathbf{E}$ is* $\mathbf{C}$-orthogonal *if $\mathbf{E}^T\mathbf{C}\mathbf{E}$ is a diagonal matrix.*

Definition 3. *A non-zero $m \times p$ matrix $\mathbf{E}$ is* $\{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_t\}$-orthogonal *if it is* $\mathbf{C}_i$-orthogonal *for all $i \in \{1, 2, \ldots, t\}$.*

Definition 4. *A non-zero $m \times p$ matrix $\mathbf{E}$ and a non-zero $m \times q$ matrix $\mathbf{F}$ are* cross $\mathbf{C}$-orthogonal *if $\mathbf{E}^T\mathbf{C}\mathbf{F} = \mathbf{0}$.*

Definition 5. *A non-zero $m \times p$ matrix $\mathbf{E}$ and a non-zero $m \times q$ matrix $\mathbf{F}$ are* cross $\{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_t\}$-orthogonal *if they are* cross $\mathbf{C}_i$-orthogonal *for all $i \in \{1, 2, \ldots, t\}$.*

Definition 6. *A non-zero vector $\mathbf{w}$ is in* $\text{span}\{\mathbf{c}_i\}_{i=1}^l$ *if there exist real numbers $\{\gamma_i\}_{i=1}^l$ such that*

$$\mathbf{w} = \sum_{i=1}^{l} \gamma_i \mathbf{c}_i.$$

We denote this as

$$\mathbf{w} \in \text{span}\{\mathbf{c}_i\}_{i=1}^l.$$

We have the following theorem about the optimal solution to the optimization problem (P_{G2}).

Theorem 2. *There exist two matrices*

$$\mathbf{T} \triangleq [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{m-r}]$$

and

$$\mathbf{W} \triangleq [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{m-r}]$$

such that $[\mathbf{S}, \mathbf{T}]$ *and* $[\mathbf{S}, \mathbf{W}]$ *are invertible and if* $r < m$ *then*

(a) $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{m-r} \in \text{span}\{\mathbf{a}_i\}_{i=1}^{p}$,

(b) $\mathbf{T}$ *is* $\big\{(\mathbf{K_X} - \mathbf{B}_2^*), (\mathbf{K_X} - \mathbf{B}_1^* - \mathbf{B}_2^*)\big\}$-orthogonal *with*

$$\mathbf{T}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{T} = \mathbf{T}^T(\mathbf{K_X} - \mathbf{B}_1^* - \mathbf{B}_2^*)\mathbf{T},$$

(c) $\mathbf{S}$ *and* $\mathbf{T}$ *are* cross $\big\{\mathbf{D}, (\mathbf{K_X} - \mathbf{B}_2^*), (\mathbf{K_X} - \mathbf{B}_1^* - \mathbf{B}_2^*)\big\}$-orthogonal,

(d) $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{m-r} \in \text{span}\{\mathbf{b}_i\}_{i=1}^{q}$,

(e) $\mathbf{W}$ *is* $\big\{\mathbf{K_Y}, (\mathbf{K_Y} - \mathbf{B}_2^*)\big\}$-orthogonal *with*

$$\mathbf{W}^T\mathbf{K_Y}\mathbf{W} = \mathbf{W}^T(\mathbf{K_Y} - \mathbf{B}_2^*)\mathbf{W}, \quad \textit{and}$$

(f) $\mathbf{S}$ *and* $\mathbf{W}$ *are* cross $\big\{\mathbf{K_Y}, (\mathbf{K_Y} - \mathbf{B}_2^*)\big\}$-orthogonal.

Proof. It suffices to consider $r < m$ case. Since $\mathbf{\Delta}^* = \mathbf{S}\mathbf{S}^T$ is rank deficient in this case, there exists $\mathbf{z}_1 \neq \mathbf{0}$ such that

$$\mathbf{S}^T\mathbf{z}_1 = \mathbf{0}.$$

Let us define

$$\mathbf{t}_1 \triangleq (\mathbf{K_X} - \mathbf{B}_2^*)^{-1}\mathbf{z}_1.$$

Therefore

$$\mathbf{S}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1 = \mathbf{0}.$$

We have from (12), (13), and (14) that

$$\frac{\mu}{2}(\mathbf{K_X} - \mathbf{B}_2^*)^{-1} = \mathbf{\Delta}^* + \mathbf{M}_1^* = \mathbf{S}\mathbf{S}^T + \sum_{i=1}^{p} \alpha_i \mathbf{a}_i \mathbf{a}_i^T.$$

On post-multiplying this by $(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1$, we obtain

$$\begin{aligned}\frac{\mu}{2}\mathbf{t}_1 &= \mathbf{S}\mathbf{S}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1 + \sum_{i=1}^{p} \alpha_i \mathbf{a}_i \mathbf{a}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1 \\ &= \sum_{i=1}^{p} \alpha_i \mathbf{a}_i \big(\mathbf{a}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1\big).\end{aligned}$$

This proves that

$$\mathbf{t}_1 \in \text{span}\{\mathbf{a}_i\}_{i=1}^{p}.$$

We next show that

$$\mathbf{t}_1 \notin \text{span}\{\mathbf{s}_i\}_{i=1}^{r}.$$

Suppose otherwise that

$$\mathbf{t}_1 \in \text{span}\{\mathbf{s}_i\}_{i=1}^{r}.$$

Then there exist real numbers $\{c_i\}_{i=1}^{r}$ such that

$$\mathbf{t}_1 = \sum_{i=1}^{r} c_i \mathbf{s}_i.$$

Since $\mathbf{S}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1 = \mathbf{0}$, we have

$$\mathbf{s}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1 = 0 \quad \text{for all} \quad i \in \{1, 2, \ldots, r\}.$$

On multiplying this by c_i and then summing over all i in $\{1, 2, \ldots, r\}$, we obtain

$$\mathbf{t}_1^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1 = 0,$$

which is a contradiction because $\mathbf{K_X} - \mathbf{B}_2^*$ is positive definite. We therefore have that

$$\mathbf{t}_1 \notin \text{span}\{\mathbf{s}_i\}_{i=1}^r.$$

We have shown so far that there exists $\mathbf{t}_1 \in \text{span}\{\mathbf{a}_i\}_{i=1}^p$ such that the rank of $[\mathbf{S}, \mathbf{t}_1]$ is $r + 1$ and

$$\mathbf{S}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1 = \mathbf{0}.$$

Let us now assume that there exists

$$\mathbf{T}_j \triangleq [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_j],$$

where

$$\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_j \in \text{span}\{\mathbf{a}_i\}_{i=1}^p$$

and $1 \leq j < m - r$ such that the rank of $[\mathbf{S}, \mathbf{T}_j]$ is $r + j$,

$$\mathbf{S}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{T}_j = \mathbf{0},$$

and

$$\mathbf{t}_k^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_l = 0$$

for all $k \neq l$ in $\{1, 2, \ldots, j\}$. Then there exists $\mathbf{z}_{j+1} \neq \mathbf{0}$ such that

$$[\mathbf{S}, \mathbf{T}_j]^T\mathbf{z}_{j+1} = \mathbf{0}.$$

Let us define

$$\mathbf{t}_{j+1} \triangleq (\mathbf{K_X} - \mathbf{B}_2^*)^{-1}\mathbf{z}_{j+1}.$$

We therefore have that

$$[\mathbf{S}, \mathbf{T}_j]^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_{j+1} = \mathbf{0}.$$

It can be shown as before that

$$\mathbf{t}_{j+1} \in \text{span}\{\mathbf{a}_i\}_{i=1}^p$$

and

$$\mathbf{t}_{j+1} \notin \text{span}\left\{\{\mathbf{s}_i\}_{i=1}^r, \{\mathbf{t}_k\}_{k=1}^j\right\}.$$

Hence, the rank of $[\mathbf{S}, \mathbf{T}_{j+1}]$, where

$$\mathbf{T}_{j+1} \triangleq [\mathbf{T}_j, \mathbf{t}_{j+1}],$$

is $r + j + 1$,

$$\mathbf{S}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{T}_{j+1} = \mathbf{0},$$

and

$$\mathbf{t}_k^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_l = 0,$$

for all $k \neq l$ in $\{1, 2, \ldots, j + 1\}$. It now follows from the mathematical induction that there exist

$$\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{m-r} \in \text{span}\{\mathbf{a}_i\}_{i=1}^p$$

such that if we define

$$\mathbf{T} \triangleq [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{m-r}],$$

then $[\mathbf{S}, \mathbf{T}]$ is invertible,

$$\mathbf{S}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{T} = \mathbf{0}, \quad \text{and}$$
$$\mathbf{T}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{T} = \mathbf{G},$$

where

$$\mathbf{G} \triangleq \text{Diag}\Big\{\big(\mathbf{t}_1^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_1\big), \big(\mathbf{t}_2^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_2\big), \ldots, \big(\mathbf{t}_{m-r}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{t}_{m-r}\big)\Big\}.$$

Since $\mathbf{B}_1^*\mathbf{T} = \mathbf{0}$ from (9) and $(\mathbf{K_X} - \mathbf{B}_1^* - \mathbf{B}_2^*)\mathbf{S} = \mathbf{DS}$ from (16), we immediately have that

$$\mathbf{S}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{T} = \mathbf{S}^T(\mathbf{K_X} - \mathbf{B}_1^* - \mathbf{B}_2^*)\mathbf{T} = \mathbf{S}^T\mathbf{DT} = \mathbf{0}, \quad \text{and}$$
$$\mathbf{T}^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{T} = \mathbf{T}^T(\mathbf{K_X} - \mathbf{B}_1^* - \mathbf{B}_2^*)\mathbf{T} = \mathbf{G}.$$

This completes the proof of parts (a) through (c) of the theorem.

For parts (d) through (f), we have from (12), (13), and (15) that

$$\frac{1}{2}(\mathbf{K_Y} - \mathbf{B}_2^*)^{-1} = \mathbf{\Delta}^* + \mathbf{M}_2^* = \mathbf{SS}^T + \sum_{i=1}^{q} \beta_i \mathbf{b}_i \mathbf{b}_i^T.$$

Similar to the previous case, we can find

$$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{m-r} \in \text{span}\{\mathbf{b}_i\}_{i=1}^q$$

such that if we define

$$\mathbf{W} \triangleq [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{m-r}],$$

then $[\mathbf{S}, \mathbf{W}]$ is invertible,

$$\mathbf{S}^T(\mathbf{K_Y} - \mathbf{B}_2^*)\mathbf{W} = \mathbf{0}, \quad \text{and}$$
$$\mathbf{W}^T(\mathbf{K_Y} - \mathbf{B}_2^*)\mathbf{W} = \mathbf{H},$$

where

$$\mathbf{H} \triangleq \text{Diag}\Big\{\big(\mathbf{w}_1^T(\mathbf{K_Y} - \mathbf{B}_2^*)\mathbf{w}_1\big), \big(\mathbf{w}_2^T(\mathbf{K_Y} - \mathbf{B}_2^*)\mathbf{w}_2\big), \ldots, \big(\mathbf{w}_{m-r}^T(\mathbf{K_Y} - \mathbf{B}_2^*)\mathbf{w}_{m-r}\big)\Big\}.$$

Since $\mathbf{B}_2^*\mathbf{W} = \mathbf{0}$ from (9), we conclude

$$\mathbf{S}^T\mathbf{K_Y}\mathbf{W} = \mathbf{S}^T(\mathbf{K_Y} - \mathbf{B}_2^*)\mathbf{W} = \mathbf{0}, \quad \text{and}$$
$$\mathbf{W}^T\mathbf{K_Y}\mathbf{W} = \mathbf{W}^T(\mathbf{K_Y} - \mathbf{B}_2^*)\mathbf{W} = \mathbf{H}.$$

This completes the proof of parts (d) through (f) of the theorem. □

We have the following corollary of Theorem 2.

Corollary 1. *If $r < m = r + p$, then we can set*

$$\mathbf{t}_i = \sqrt{\alpha_i}\mathbf{a}_i$$

for all i in $\{1, 2, \ldots, p\}$. Similarly, if $r < m = r + q$, then we can set

$$\mathbf{w}_i = \sqrt{\beta_i}\mathbf{b}_i$$

for all i in $\{1, 2, \ldots, q\}$.

Proof. Let $r < m = r + p$ and let us set

$$\mathbf{t}_i = \sqrt{\alpha_i}\mathbf{a}_i$$

for all i in $\{1, 2, \ldots, p\}$ in the definition of $\mathbf{T}$. We have from (12), (13), and (14) that

$$\frac{\mu}{2}(\mathbf{K_X} - \mathbf{B}_2^*)^{-1} = \sum_{i=1}^{r} \lambda_i \mathbf{s}_i \mathbf{s}_i^T + \sum_{i=1}^{p} \alpha_i \mathbf{a}_i \mathbf{a}_i^T. \tag{17}$$

Now, on post-multiplying (17) by $(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_1$, we obtain

$$\frac{\mu}{2}\mathbf{s}_1 = \sum_{i=1}^{r} \lambda_i \mathbf{s}_i \left(\mathbf{s}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_1\right) + \sum_{i=1}^{p} \alpha_i \mathbf{a}_i \left(\mathbf{a}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_1\right),$$

which can be re-written as

$$\mathbf{s}_1\left(\frac{\mu}{2} - \lambda_1\left(\mathbf{s}_1^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_1\right)\right) - \sum_{i=2}^{r} \lambda_i \mathbf{s}_i \left(\mathbf{s}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_1\right) = \sum_{i=1}^{p} \alpha_i \mathbf{a}_i \left(\mathbf{a}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_1\right). \tag{18}$$

Since $[\mathbf{S}, \mathbf{T}]$ is invertible from (17), its columns are linearly independent. Hence, the coefficients of all vectors in (18) must be zero. Therefore,

$$\begin{aligned}
\lambda_1 \mathbf{s}_1^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_1 &= \frac{\mu}{2}, \\
\mathbf{s}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_1 &= 0, \quad \forall i \in \{2, \ldots, r\}, \quad \text{and} \\
\mathbf{a}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_1 &= 0, \quad \forall i \in \{1, \ldots, p\}.
\end{aligned}$$

Likewise, on post-multiplying (17) by $(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_2, \ldots, (\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_r, (\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{a}_1, \ldots, (\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{a}_p$ and then equating all coefficients to zero, we obtain similar equations. In summary,

$$\begin{aligned}
\lambda_i \mathbf{s}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_i &= \frac{\mu}{2}, \quad \forall i \in \{1, \ldots, r\}, \\
\alpha_i \mathbf{a}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{a}_i &= \frac{\mu}{2}, \quad \forall i \in \{1, \ldots, p\}, \\
\mathbf{s}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{s}_j &= 0, \quad \forall i, j \in \{1, \ldots, r\}, i \neq j, \\
\mathbf{a}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{a}_j &= 0, \quad \forall i, j \in \{1, \ldots, p\}, i \neq j, \quad \text{and} \\
\mathbf{s}_i^T(\mathbf{K_X} - \mathbf{B}_2^*)\mathbf{a}_j &= 0, \quad \forall i \in \{1, \ldots, r\}, \forall j \in \{1, \ldots, p\}.
\end{aligned}$$

Hence,

$$[\mathbf{S}, \mathbf{T}]^T(\mathbf{K_X} - \mathbf{B}_2^*)[\mathbf{S}, \mathbf{T}] = \frac{\mu}{2}\mathbf{I}_m. \tag{19}$$

Parts (a) through (c) of Theorem 2 follow immediately from (9), (10), and (19) because $\mathbf{M}_1^* = \mathbf{T}\mathbf{T}^T$ in this case.

The proof for the case when $r < m = r + q$ is exactly similar. It starts with the following from (12), (13), and (15)

$$\frac{1}{2}(\mathbf{K_Y} - \mathbf{B}_2^*)^{-1} = \mathbf{\Delta}^* + \mathbf{M}_2^* = \sum_{i=1}^{r} \lambda_i \mathbf{s}_i \mathbf{s}_i^T + \sum_{i=1}^{q} \beta_i \mathbf{b}_i \mathbf{b}_i^T.$$

□

In summary, the key properties of the optimal Gaussian solution are as follows. If $\mathbf{\Delta}^*$ (and hence $\mathbf{S}$) is not invertible, then there exist two matrices $\mathbf{T}$ and $\mathbf{W}$ such that their columns respectively are in span$\{\mathbf{a}_i\}_{i=1}^p$ and span$\{\mathbf{b}_i\}_{i=1}^q$, $[\mathbf{S}, \mathbf{T}]$ and $[\mathbf{S}, \mathbf{W}]$ are invertible, $\mathbf{S}$ and $\mathbf{T}$ are cross $(\mathbf{K_X} - \mathbf{B}_2^*)$-orthogonal, and $\mathbf{S}$ and $\mathbf{W}$ are cross $(\mathbf{K_Y} - \mathbf{B}_2^*)$-orthogonal. We shall exploit these properties in the next section to prove the optimality of an optimization problem, which is central to prove our main result.

6 Converse Ingredients

Let us define an optimization problem as

$$\begin{aligned}
(P) \quad \triangleq \quad & \min_{\mathbf{U}, \mathbf{V}} \; \mu I(\mathbf{X}; \mathbf{U}|\mathbf{V}) + I(\mathbf{Y}; \mathbf{V}) \\
& \text{subject to} \;\; \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} \preccurlyeq \mathbf{D} \quad \text{and} \\
& \qquad \mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{V},
\end{aligned}$$

where $\mathbf{X}, \mathbf{Y}, \mathbf{D}$, and μ are defined as before. We refer to this problem as the *main optimization problem* and denote it by (P). We have the following theorem.

Theorem 3. *A Gaussian* $(\mathbf{U}, \mathbf{V})$ *is an optimal solution of the main optimization problem* (P).

We prove this theorem in the remainder of the section. The proof for μ in $[0,1]$ is easy. In this case, the objective of (P) can be lower bounded as

$$\begin{aligned}
\mu I(\mathbf{X};\mathbf{U}|\mathbf{V}) + I(\mathbf{Y};\mathbf{V}) &= \mu I(\mathbf{X};\mathbf{U},\mathbf{V}) - \mu I(\mathbf{X};\mathbf{V}) + I(\mathbf{Y};\mathbf{V}) \\
&= \mu I(\mathbf{X};\mathbf{U}) + \mu I(\mathbf{X};\mathbf{V}|\mathbf{U}) + \mu[I(\mathbf{Y};\mathbf{V}) - I(\mathbf{X};\mathbf{V})] + (1-\mu) I(\mathbf{Y};\mathbf{V}) \\
&\geq \mu I(\mathbf{X};\mathbf{U}) \qquad (20) \\
&= \mu h(\mathbf{X}) - \mu h(\mathbf{X}|\mathbf{U}) \\
&\geq \frac{\mu}{2} \log \frac{|\mathbf{K}_{\mathbf{X}}|}{|\mathbf{K}_{\mathbf{X}|\mathbf{U}}|}, \qquad (21)
\end{aligned}$$

where

(20) follows because of the facts that

$$I(\mathbf{Y};\mathbf{V}) \geq 0$$

and

$$I(\mathbf{X};\mathbf{V}|\mathbf{U}) \geq 0,$$

and we have

$$I(\mathbf{Y};\mathbf{V}) - I(\mathbf{X};\mathbf{V}) \geq 0$$

because of the data processing inequality [21, Theorem 2.8.1] and the Markov chain $\mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{V}$, and

(21) follows because the Gaussian distribution maximizes the differential entropy for a given covariance matrix [21, Theorem 8.6.5], i.e.,

$$h(\mathbf{X}|\mathbf{U}) \leq \frac{1}{2} \log \left((2\pi e)^m \left| \mathbf{K}_{\mathbf{X}|\mathbf{U}} \right| \right).$$

Inequalities (20) and (21) become equalities if we choose a Gaussian $(\mathbf{U}, \mathbf{V})$ such that $\mathbf{V}$ is independent of $(\mathbf{X}, \mathbf{Y}, \mathbf{U})$. Because of the distortion constraint in (P), the conditional covariance of $\mathbf{X}$ given $(\mathbf{U}, \mathbf{V})$ should satisfy

$$\mathbf{0} \preccurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} = \mathbf{K}_{\mathbf{X}|\mathbf{U}} \preccurlyeq \mathbf{D}.$$

Since conditioning reduces covariance in a positive semidefinite sense, we also have

$$\mathbf{K}_{\mathbf{X}|\mathbf{U}} \preccurlyeq \mathbf{K}_{\mathbf{X}}.$$

Hence, if μ is in $[0,1]$, then a Gaussian $(\mathbf{U}, \mathbf{V})$ is an optimal solution of the *main optimization problem* (P) and the optimal value is

$$\begin{aligned}
v(P) = \quad & \min_{\mathbf{K}_{\mathbf{X}|\mathbf{U}}} \quad \frac{\mu}{2} \log \frac{|\mathbf{K}_{\mathbf{X}}|}{|\mathbf{K}_{\mathbf{X}|\mathbf{U}}|} \\
& \text{subject to} \quad \mathbf{K}_{\mathbf{X}} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U}} \succcurlyeq \mathbf{0} \quad \text{and} \\
& \qquad\qquad\quad \mathbf{D} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U}} \\
= \; & v(P_{pt-pt}). \qquad (22)
\end{aligned}$$

We therefore assume that $\mu > 1$ in the rest of the section.

Let us first restrict the solution space of (P) to Gaussian distributions. This results in an optimization problem (P_{G1}), or equivalently (P_{G2}), defined in Section 5. For convenience, we shall work with the (P_{G2}) formulation. First note that since restricting the solution space to Gaussian distributions can only increase the optimal value of the *main optimization problem* (P), we immediately have

$$v(P_{G1}) = v(P_{G2}) \geq v(P). \qquad (23)$$

So, it suffices to prove the reverse inequality

$$v(P_{G2}) \leq v(P).$$

Let $(\mathbf{B}_1^*, \mathbf{B}_2^*)$ be an optimal solution to (P_{G2}). As discussed in Section 5, $(\mathbf{B}_1^*, \mathbf{B}_2^*)$ gives three matrices $\mathbf{S}, \mathbf{T}$, and $\mathbf{W}$ that satisfy the properties in Theorem 2. Using these properties, the optimal value of (P_{G2}) can be expressed as

$$
\begin{aligned}
v\left(P_{G2}\right) &= \frac{\mu}{2}\log\frac{|\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*|}{|\mathbf{K}_\mathbf{X}-\mathbf{B}_1^*-\mathbf{B}_2^*|}+\frac{1}{2}\log\frac{|\mathbf{K}_\mathbf{Y}|}{|\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*|} \\
&= \frac{\mu}{2}\log\frac{\left|[\mathbf{S},\mathbf{T}]^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)[\mathbf{S},\mathbf{T}]\right|}{\left|[\mathbf{S},\mathbf{T}]^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_1^*-\mathbf{B}_2^*\right)[\mathbf{S},\mathbf{T}]\right|}+\frac{1}{2}\log\frac{\left|[\mathbf{S},\mathbf{W}]^T\mathbf{K}_\mathbf{Y}[\mathbf{S},\mathbf{W}]\right|}{\left|[\mathbf{S},\mathbf{W}]^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)[\mathbf{S},\mathbf{W}]\right|} \qquad (24)\\
&= \frac{\mu}{2}\log\frac{\left|\begin{pmatrix}\mathbf{S}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{S} & \mathbf{0}\\ \mathbf{0} & \mathbf{T}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{T}\end{pmatrix}\right|}{\left|\begin{pmatrix}\mathbf{S}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_1^*-\mathbf{B}_2^*\right)\mathbf{S} & \mathbf{0}\\ \mathbf{0} & \mathbf{T}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_1^*-\mathbf{B}_2^*\right)\mathbf{T}\end{pmatrix}\right|} \\
&\quad +\frac{1}{2}\log\frac{\left|\begin{pmatrix}\mathbf{S}^T\mathbf{K}_\mathbf{Y}\mathbf{S} & \mathbf{0}\\ \mathbf{0} & \mathbf{W}^T\mathbf{K}_\mathbf{Y}\mathbf{W}\end{pmatrix}\right|}{\left|\begin{pmatrix}\mathbf{S}^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{S} & \mathbf{0}\\ \mathbf{0} & \mathbf{W}^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{W}\end{pmatrix}\right|} \qquad (25)\\
&= \frac{\mu}{2}\log\frac{\left|\mathbf{S}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{S}\right|\left|\mathbf{T}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{T}\right|}{\left|\mathbf{S}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_1^*-\mathbf{B}_2^*\right)\mathbf{S}\right|\left|\mathbf{T}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_1^*-\mathbf{B}_2^*\right)\mathbf{T}\right|} \\
&\quad +\frac{1}{2}\log\frac{\left|\mathbf{S}^T\mathbf{K}_\mathbf{Y}\mathbf{S}\right|\left|\mathbf{W}^T\mathbf{K}_\mathbf{Y}\mathbf{W}\right|}{\left|\mathbf{S}^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{S}\right|\left|\mathbf{W}^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{W}\right|} \\
&= \frac{\mu}{2}\log\frac{\left|\mathbf{S}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{S}\right|}{\left|\mathbf{S}^T\mathbf{D}\mathbf{S}\right|}+\frac{1}{2}\log\frac{\left|\mathbf{S}^T\mathbf{K}_\mathbf{Y}\mathbf{S}\right|}{\left|\mathbf{S}^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{S}\right|}, \qquad (26)
\end{aligned}
$$

where

(24) follows because $[\mathbf{S},\mathbf{T}]$ and $[\mathbf{S},\mathbf{W}]$ are invertible,

(25) follows because $\mathbf{S}$ and $\mathbf{T}$ are cross $\left\{(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*),(\mathbf{K}_\mathbf{X}-\mathbf{B}_1^*-\mathbf{B}_2^*)\right\}$-orthogonal, and $\mathbf{S}$ and $\mathbf{W}$ are cross $\left\{\mathbf{K}_\mathbf{Y},(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*)\right\}$-orthogonal, and

(26) follows from (16) and the facts that

$$
\begin{aligned}
\mathbf{T}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{T} &= \mathbf{T}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_1^*-\mathbf{B}_2^*\right)\mathbf{T} \quad \text{and}\\
\mathbf{W}^T\mathbf{K}_\mathbf{Y}\mathbf{W} &= \mathbf{W}^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{W}.
\end{aligned}
$$

6.1 Distortion Projection

The special structure to the optimal Gaussian solution of (P_{G2}) suggests a way to lower bound (P) by projecting the sources $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbf{S}$ and imposing the distortion constraint on the subspace spanned by the columns of $\mathbf{S}$. Note that the distortion constraint is tight on this subspace for the optimal Gaussian solution. We refer to this method of lower bounding (P) as *distortion projection*. Let us define

$$
\begin{aligned}
\tilde{\mathbf{X}} &\triangleq \mathbf{S}^T\mathbf{X},\\
\tilde{\mathbf{Y}} &\triangleq \mathbf{S}^T\mathbf{Y},\\
\tilde{\mathbf{D}} &\triangleq \mathbf{S}^T\mathbf{D}\mathbf{S},\\
\tilde{\mathbf{B}}_1^* &\triangleq \mathbf{S}^T\mathbf{B}_1^*\mathbf{S},\\
\tilde{\mathbf{B}}_2^* &\triangleq \mathbf{S}^T\mathbf{B}_2^*\mathbf{S},\\
\tilde{\mathbf{M}}_1^* &\triangleq \left(\mathbf{S}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{S}\right)^{-1}\mathbf{S}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{M}_1^*\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{S}\left(\mathbf{S}^T\left(\mathbf{K}_\mathbf{X}-\mathbf{B}_2^*\right)\mathbf{S}\right)^{-1}, \quad \text{and}\\
\tilde{\mathbf{M}}_2^* &\triangleq \left(\mathbf{S}^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{S}\right)^{-1}\mathbf{S}^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{M}_2^*\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{S}\left(\mathbf{S}^T\left(\mathbf{K}_\mathbf{Y}-\mathbf{B}_2^*\right)\mathbf{S}\right)^{-1}.
\end{aligned}
$$

Since $\mathbf{S}$ has full column rank, we immediately have that

$$
\begin{aligned}
\mathbf{K}_{\tilde{\mathbf{X}}},\mathbf{K}_{\tilde{\mathbf{Y}}},\tilde{\mathbf{D}} &\succ \mathbf{0},\\
\tilde{\mathbf{B}}_1^*,\tilde{\mathbf{B}}_2^* &\succcurlyeq \mathbf{0}, \quad \text{and}\\
\tilde{\mathbf{M}}_1^*,\tilde{\mathbf{M}}_2^* &\succcurlyeq \mathbf{0}.
\end{aligned}
$$

The *projected optimization problem* $(\tilde{P})$ is now defined as

$$\begin{aligned}(\tilde{P}) \qquad \triangleq \qquad & \min_{\mathbf{U},\mathbf{V}} \quad \mu I(\tilde{\mathbf{X}};\mathbf{U}|\mathbf{V}) + I(\tilde{\mathbf{Y}};\mathbf{V}) \\ & \text{subject to} \quad \mathbf{K}_{\tilde{\mathbf{X}}|\mathbf{U},\mathbf{V}} \preccurlyeq \tilde{\mathbf{D}} \quad \text{and} \\ & \qquad\qquad \tilde{\mathbf{X}} \leftrightarrow \tilde{\mathbf{Y}} \leftrightarrow \mathbf{V}.\end{aligned}$$

We next show that the *main optimization problem* (P) is lower bounded by the *projected optimization problem* $(\tilde{P})$. Since $[\mathbf{S},\mathbf{T}]$ and $[\mathbf{S},\mathbf{W}]$ are invertible and mutual information is nonnegative, we have

$$\begin{aligned}\mu I(\mathbf{X};\mathbf{U}|\mathbf{V}) + I(\mathbf{Y};\mathbf{V}) &= \mu I\left(\mathbf{S}^T\mathbf{X},\mathbf{T}^T\mathbf{X};\mathbf{U}|\mathbf{V}\right) + I\left(\mathbf{S}^T\mathbf{Y},\mathbf{W}^T\mathbf{Y};\mathbf{V}\right) \\ &= \mu I\left(\mathbf{S}^T\mathbf{X};\mathbf{U}|\mathbf{V}\right) + \mu I\left(\mathbf{T}^T\mathbf{X};\mathbf{U}|\mathbf{V},\mathbf{S}^T\mathbf{X}\right) + I\left(\mathbf{S}^T\mathbf{Y};\mathbf{V}\right) + I\left(\mathbf{W}^T\mathbf{Y};\mathbf{V}|\mathbf{S}^T\mathbf{Y}\right) \\ &\geq \mu I\big(\tilde{\mathbf{X}};\mathbf{U}|\mathbf{V}\big) + I\big(\tilde{\mathbf{Y}};\mathbf{V}\big). \qquad (27)\end{aligned}$$

Consider any $(\mathbf{U},\mathbf{V})$ feasible for (P). Then

$$\mathbf{D} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} \quad \text{and} \qquad (28)$$

$$\mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{V} \qquad (29)$$

Now (28) implies

$$\tilde{\mathbf{D}} = \mathbf{S}^T\mathbf{D}\mathbf{S} \succcurlyeq \mathbf{S}^T\mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}}\mathbf{S} = \mathbf{K}_{\tilde{\mathbf{X}}|\mathbf{U},\mathbf{V}}, \qquad (30)$$

and (29) yields

$$\begin{aligned}0 &= I\left(\mathbf{X};\mathbf{V}|\mathbf{Y}\right) \\ &= I\left(\mathbf{S}^T\mathbf{X};\mathbf{V}|\mathbf{Y}\right) + I\left(\mathbf{T}^T\mathbf{X};\mathbf{V}|\mathbf{Y},\mathbf{S}^T\mathbf{X}\right) \\ &\geq I\left(\mathbf{S}^T\mathbf{X};\mathbf{V}|\mathbf{Y}\right) \qquad (31) \\ &= I\left(\mathbf{S}^T\mathbf{X};\mathbf{V}|\mathbf{S}^T\mathbf{Y},\mathbf{W}^T\mathbf{Y}\right) \qquad (32) \\ &= h\left(\mathbf{S}^T\mathbf{X}|\mathbf{S}^T\mathbf{Y},\mathbf{W}^T\mathbf{Y}\right) - h\left(\mathbf{S}^T\mathbf{X}|\mathbf{V},\mathbf{S}^T\mathbf{Y},\mathbf{W}^T\mathbf{Y}\right) \\ &\geq h\left(\mathbf{S}^T\mathbf{X}|\mathbf{S}^T\mathbf{Y}\right) - h\left(\mathbf{S}^T\mathbf{X}|\mathbf{V},\mathbf{S}^T\mathbf{Y}\right) \qquad (33) \\ &= I\left(\mathbf{S}^T\mathbf{X};\mathbf{V}|\mathbf{S}^T\mathbf{Y}\right) \\ &= I\big(\tilde{\mathbf{X}};\mathbf{V}|\tilde{\mathbf{Y}}\big) \\ &\geq 0, \qquad (34)\end{aligned}$$

where

(31) and (34) follows because mutual information is nonnegative,

(32) follows because $[\mathbf{S},\mathbf{W}]$ is invertible, and

(33) follows because conditioning reduces entropy and we have from Theorem 2 that $\mathbf{W}^T\mathbf{Y}$ is independent of $\mathbf{S}^T\mathbf{Y}$, which implies that $\mathbf{W}^T\mathbf{Y}$ is also independent of $\mathbf{S}^T\mathbf{X}$ because $\mathbf{X} = \mathbf{Y} + \mathbf{N}$ by assumption.

Now (34) is equivalent to

$$\tilde{\mathbf{X}} \leftrightarrow \tilde{\mathbf{Y}} \leftrightarrow \mathbf{V},$$

which together with (30) implies that $(\mathbf{U},\mathbf{V})$ is feasible for $(\tilde{P})$. Hence, the feasible set of (P) is contained in that of $(\tilde{P})$. Moreover, (27) above implies that the objective of (P) is no less than that of $(\tilde{P})$. We therefore have that the *projected optimization problem* $(\tilde{P})$ lower bounds the *main optimization problem* (P), i.e.,

$$v\left(P\right) \geq v(\tilde{P}). \qquad (35)$$

By restricting the solution space of $(\tilde{P})$ to Gaussian distributions, we obtain its Gaussian version

$$
\begin{aligned}
(\tilde{P}_{G2}) \quad \triangleq \quad \min_{\tilde{\mathbf{B}}_1,\tilde{\mathbf{B}}_2} \quad & \frac{\mu}{2}\log\frac{|\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_2|}{|\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_1-\tilde{\mathbf{B}}_2|}+\frac{1}{2}\log\frac{|\mathbf{K}_{\tilde{\mathbf{Y}}}|}{|\mathbf{K}_{\tilde{\mathbf{Y}}}-\tilde{\mathbf{B}}_2|} \\
\text{subject to} \quad & \tilde{\mathbf{B}}_i \succcurlyeq \mathbf{0} \ \text{ for all } \ i\in\{1,2\}, \ \text{ and} \\
& \tilde{\mathbf{D}} \succcurlyeq \mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_1-\tilde{\mathbf{B}}_2.
\end{aligned}
$$

It is easy to verify that the projected optimal Gaussian solution $(\tilde{\mathbf{B}}_1^*,\tilde{\mathbf{B}}_2^*)$ is feasible for $(\tilde{P}_{G2})$ and it meets the projected distortion constraint $\tilde{\mathbf{D}}$ with equality from (16). We next show that $(\tilde{\mathbf{B}}_1^*,\tilde{\mathbf{B}}_2^*)$ is in fact optimal for $(\tilde{P})$.

Remark 1: If $r=m$, then there is no need for *distortion projection* because $\mathbf{S}$ is invertible, and hence so is $\mathbf{\Delta}^*$.

6.2 Source Enhancement

In this subsection, we use the KKT conditions (7) through (11) satisfied by $(\mathbf{B}_1^*,\mathbf{B}_2^*)$ to derive conditions that must be satisfied by $(\tilde{\mathbf{B}}_1^*,\tilde{\mathbf{B}}_2^*)$. These conditions are then used to define the *enhanced optimization problem,* which lower bounds $(\tilde{P})$. We show that the optimal solution to the *enhanced optimization problem* is Gaussian, in particular $(\tilde{\mathbf{B}}_1^*,\tilde{\mathbf{B}}_2^*)$ is optimal for the problem. This will in turn prove that $(\tilde{\mathbf{B}}_1^*,\tilde{\mathbf{B}}_2^*)$ is optimal for $(\tilde{P})$. This approach of lower bounding is referred to as the *source enhancement* [6] and is similar to the *channel enhancement* idea of Weingarten *et al.* [3].

We start with the following key lemma.

Lemma 4. *For $\mathbf{K}_{\tilde{\mathbf{X}}},\mathbf{K}_{\tilde{\mathbf{Y}}},\tilde{\mathbf{D}},\tilde{\mathbf{B}}_i^*$, and $\tilde{\mathbf{M}}_i^*$, where $i=1,2$, defined as above, the following hold*

$$\mathbf{I}_r=\frac{\mu}{2}\big(\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*\big)^{-1}-\tilde{\mathbf{M}}_1^*=\frac{1}{2}\big(\mathbf{K}_{\tilde{\mathbf{Y}}}-\tilde{\mathbf{B}}_2^*\big)^{-1}-\tilde{\mathbf{M}}_2^*, \tag{36}$$

$$\tilde{\mathbf{B}}_i^*\tilde{\mathbf{M}}_i^*=\mathbf{0} \ \ \textit{for all} \ \ i\in\{1,2\}, \quad \textit{and} \tag{37}$$

$$\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_1^*-\tilde{\mathbf{B}}_2^*=\tilde{\mathbf{D}}. \tag{38}$$

Proof. See Appendix D. ☐

Let $\mathbf{K}_{\hat{\mathbf{X}}}$ and $\mathbf{K}_{\hat{\mathbf{Y}}}$ be two real symmetric matrices satisfying

$$\frac{\mu}{2}\big(\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*\big)^{-1}-\tilde{\mathbf{M}}_1^*=\frac{\mu}{2}\big(\mathbf{K}_{\hat{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*\big)^{-1} \quad \text{and} \tag{39}$$

$$\frac{1}{2}\big(\mathbf{K}_{\tilde{\mathbf{Y}}}-\tilde{\mathbf{B}}_2^*\big)^{-1}-\tilde{\mathbf{M}}_2^*=\frac{1}{2}\big(\mathbf{K}_{\hat{\mathbf{Y}}}-\tilde{\mathbf{B}}_2^*\big)^{-1}. \tag{40}$$

We now have the following lemma, which is similar to [3, Lemmas 11, 12].

Lemma 5. *For $\mathbf{K}_{\tilde{\mathbf{X}}},\mathbf{K}_{\tilde{\mathbf{Y}}},\mathbf{K}_{\hat{\mathbf{X}}},\mathbf{K}_{\hat{\mathbf{Y}}},\tilde{\mathbf{B}}_i^*,\tilde{\mathbf{M}}_i^*$, $i=1,2$, defined as above, and $\mu>1$, the following hold*

$$\mathbf{K}_{\hat{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*=\frac{\mu}{2}\mathbf{I}_r, \tag{41}$$

$$\mathbf{K}_{\hat{\mathbf{Y}}}-\tilde{\mathbf{B}}_2^*=\frac{1}{2}\mathbf{I}_r, \tag{42}$$

$$\mathbf{K}_{\hat{\mathbf{X}}}\succ\mathbf{K}_{\hat{\mathbf{Y}}}\succcurlyeq\mathbf{K}_{\tilde{\mathbf{Y}}}\succ\mathbf{0}, \tag{43}$$

$$\mathbf{K}_{\hat{\mathbf{X}}}\succcurlyeq\mathbf{K}_{\tilde{\mathbf{X}}}\succ\mathbf{0}, \tag{44}$$

$$\frac{|\mathbf{K}_{\tilde{\mathbf{Y}}}|}{|\mathbf{K}_{\tilde{\mathbf{Y}}}-\tilde{\mathbf{B}}_2^*|}=\frac{|\mathbf{K}_{\hat{\mathbf{Y}}}|}{|\mathbf{K}_{\hat{\mathbf{Y}}}-\tilde{\mathbf{B}}_2^*|}, \quad \textit{and} \tag{45}$$

$$\frac{|\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*|}{|\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_1^*-\tilde{\mathbf{B}}_2^*|}=\frac{|\mathbf{K}_{\hat{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*|}{|\mathbf{K}_{\hat{\mathbf{X}}}-\tilde{\mathbf{B}}_1^*-\tilde{\mathbf{B}}_2^*|}. \tag{46}$$

Proof. See Appendix E. ☐

Let $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ be two zero-mean Gaussian random vectors with covariance matrices $\mathbf{K}_{\hat{\mathbf{X}}}$ and $\mathbf{K}_{\hat{\mathbf{Y}}}$, respectively. Since $\mathbf{K}_{\hat{\mathbf{X}}}\succ\mathbf{K}_{\hat{\mathbf{Y}}}$ from (43), we can write

$$\hat{\mathbf{X}}=\hat{\mathbf{Y}}+\hat{\mathbf{N}},$$

where $\hat{\mathbf{N}}$ is a zero-mean Gaussian random vector with the covariance matrix

$$\mathbf{K}_{\hat{\mathbf{N}}}=\mathbf{K}_{\hat{\mathbf{X}}}-\mathbf{K}_{\hat{\mathbf{Y}}}=\frac{\mu-1}{2}\mathbf{I}_r$$

and is independent of $\hat{\mathbf{Y}}$. Similarly, we can use (43) and (44) to relate $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ with $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$, respectively, and write

$$\begin{aligned}\hat{\mathbf{X}} &= \tilde{\mathbf{X}} + \mathbf{N}_1 \quad \text{and} \\ \hat{\mathbf{Y}} &= \tilde{\mathbf{Y}} + \mathbf{N}_2,\end{aligned}$$

where $\mathbf{N}_1$ and $\mathbf{N}_2$ are two zero-mean Gaussian random vectors with covariance matrices

$$\begin{aligned}\mathbf{K}_{\mathbf{N}_1} &= \mathbf{K}_{\hat{\mathbf{X}}} - \mathbf{K}_{\tilde{\mathbf{X}}} \quad \text{and} \\ \mathbf{K}_{\mathbf{N}_2} &= \mathbf{K}_{\hat{\mathbf{Y}}} - \mathbf{K}_{\tilde{\mathbf{Y}}},\end{aligned}$$

respectively, and they are independent of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$. Using (38), we define

$$\hat{\mathbf{D}} \triangleq \tilde{\mathbf{D}} + \mathbf{K}_{\mathbf{N}_1} = \mathbf{K}_{\hat{\mathbf{X}}} - \tilde{\mathbf{B}}_1^* - \tilde{\mathbf{B}}_2^*. \tag{47}$$

The *enhanced optimization problem* $(\hat{P})$ is now defined as

$$\begin{aligned}(\hat{P}) \qquad \triangleq \qquad & \min_{\mathbf{U},\mathbf{V}} \ \mu I(\hat{\mathbf{X}};\mathbf{U}|\mathbf{V}) + I(\hat{\mathbf{Y}};\mathbf{V}) \\ & \text{subject to} \ \ \mathbf{K}_{\hat{\mathbf{X}}|\mathbf{U},\mathbf{V}} \preccurlyeq \hat{\mathbf{D}} \quad \text{and} \\ & \qquad \hat{\mathbf{X}} \leftrightarrow \hat{\mathbf{Y}} \leftrightarrow \mathbf{V}.\end{aligned}$$

We next show that $(\hat{P})$ lower bounds $(\tilde{P})$. Consider any $(\mathbf{U},\mathbf{V})$ feasible for $(\tilde{P})$. Without loss of optimality, we can assume that the joint distribution between $\tilde{\mathbf{X}}$, $\tilde{\mathbf{Y}}$, $\mathbf{U}$, and $\mathbf{V}$ is

$$\tilde{p} \triangleq p_{\tilde{\mathbf{X}},\tilde{\mathbf{Y}}} p_{\mathbf{U}|\tilde{\mathbf{X}},\mathbf{V}} p_{\mathbf{V}|\tilde{\mathbf{Y}}}.$$

Now, $\tilde{p}$ induces two conditional distributions as follows

$$\begin{aligned}p_{\mathbf{V}|\hat{\mathbf{Y}}} &= \int_{\tilde{\mathbf{Y}}} p_{\mathbf{V}|\tilde{\mathbf{Y}}} p_{\tilde{\mathbf{Y}}|\hat{\mathbf{Y}}} \\ p_{\mathbf{U}|\hat{\mathbf{X}},\mathbf{V}} &= \int_{\tilde{\mathbf{X}}} p_{\mathbf{U}|\tilde{\mathbf{X}},\mathbf{V}} p_{\tilde{\mathbf{X}}|\hat{\mathbf{X}},\mathbf{V}},\end{aligned}$$

where

$$p_{\tilde{\mathbf{X}}|\hat{\mathbf{X}},\mathbf{V}} = \frac{p_{\tilde{\mathbf{X}},\hat{\mathbf{X}}} p_{\mathbf{V}|\tilde{\mathbf{X}}}}{\int_{\tilde{\mathbf{X}}} p_{\tilde{\mathbf{X}},\hat{\mathbf{X}}} p_{\mathbf{V}|\tilde{\mathbf{X}}}}.$$

Then

$$\hat{p} \triangleq p_{\hat{\mathbf{X}},\hat{\mathbf{Y}}} p_{\mathbf{U}|\hat{\mathbf{X}},\mathbf{V}} p_{\mathbf{V}|\hat{\mathbf{Y}}}$$

is a joint distribution between $\hat{\mathbf{X}}$, $\hat{\mathbf{Y}}$, $\mathbf{U}$, and $\mathbf{V}$. It is clear that $\hat{p}$ satisfies the Markov condition

$$\hat{\mathbf{X}} \leftrightarrow \hat{\mathbf{Y}} \leftrightarrow \mathbf{V}. \tag{48}$$

Moreover, (47) and the distortion constraint in the definition of $(\tilde{P})$ yield

$$\mathbf{K}_{\hat{\mathbf{X}}|\mathbf{U},\mathbf{V}} = \mathbf{K}_{\tilde{\mathbf{X}}|\mathbf{U},\mathbf{V}} + \mathbf{K}_{\mathbf{N}_1} \preccurlyeq \tilde{\mathbf{D}} + \mathbf{K}_{\mathbf{N}_1} = \hat{\mathbf{D}}. \tag{49}$$

We next use the chain rule of mutual information to obtain

$$\begin{aligned}I(\tilde{\mathbf{X}},\hat{\mathbf{X}};\mathbf{U}|\mathbf{V}) &= I(\hat{\mathbf{X}};\mathbf{U}|\mathbf{V}) + I(\tilde{\mathbf{X}};\mathbf{U}|\mathbf{V},\hat{\mathbf{X}}) \\ &= I(\tilde{\mathbf{X}};\mathbf{U}|\mathbf{V}) + I(\hat{\mathbf{X}};\mathbf{U}|\mathbf{V},\tilde{\mathbf{X}}) \\ &= I(\tilde{\mathbf{X}};\mathbf{U}|\mathbf{V})\end{aligned}$$

and

$$\begin{aligned}I(\tilde{\mathbf{Y}},\hat{\mathbf{Y}};\mathbf{V}) &= I(\hat{\mathbf{Y}};\mathbf{V}) + I(\tilde{\mathbf{Y}};\mathbf{V}|\hat{\mathbf{Y}}) \\ &= I(\tilde{\mathbf{Y}};\mathbf{V}) + I(\hat{\mathbf{Y}};\mathbf{V}|\tilde{\mathbf{Y}}) \\ &= I(\tilde{\mathbf{Y}};\mathbf{V}).\end{aligned}$$

Since mutual information is nonnegative, these imply that

$$I(\tilde{\mathbf{X}};\mathbf{U}|\mathbf{V}) \geq I(\hat{\mathbf{X}};\mathbf{U}|\mathbf{V}) \tag{50}$$

and

$$I(\tilde{\mathbf{Y}};\mathbf{V}) \geq I(\hat{\mathbf{Y}};\mathbf{V}) \tag{51}$$

Now (48) and (49) together imply that the distribution $\hat{p}$, and hence $(\mathbf{U},\mathbf{V})$, is feasible for $(\hat{P})$. Therefore, the feasible set of $(\tilde{P})$ is contained in that of $(\hat{P})$. Moreover, (50) and (51) assert that the objective value of $(\hat{P})$ is no more than that of $(\tilde{P})$. We therefore conclude that the *enhanced optimization problem* $(\hat{P})$ lower bounds the *projected optimization problem* $(\tilde{P})$, i.e.,

$$v(\tilde{P}) \geq v(\hat{P}). \tag{52}$$

Remark 2: If $r < m = r + p$, then there is no need to enhance the source $\tilde{\mathbf{X}}$ and the distortion $\tilde{\mathbf{D}}$ because $\mathbf{M}_1^* = \mathbf{T}\mathbf{T}^T$ from Corollary 1, and hence $\tilde{\mathbf{M}}_1^* = \mathbf{0}$. Similarly, if $r < m = r + q$, then there is no need to enhance the source $\tilde{\mathbf{Y}}$ because $\mathbf{M}_2^* = \mathbf{W}\mathbf{W}^T$ from Corollary 1 again, and hence $\tilde{\mathbf{M}}_2^* = \mathbf{0}$. Finally, if $r < m = r + p = r + q$, then there is no need for *source enhancement*.

6.3 Oohama's Approach

We now apply Oohama's approach [1] to prove that $(\tilde{\mathbf{B}}_1^*, \tilde{\mathbf{B}}_2^*)$ is optimal for $(\hat{P})$. The objective of $(\hat{P})$ can be decomposed as

$$\mu I\big(\hat{\mathbf{X}};\mathbf{U}|\mathbf{V}\big) + I\big(\hat{\mathbf{Y}};\mathbf{V}\big) = \mu I\big(\hat{\mathbf{X}};\mathbf{U},\mathbf{V}\big) - \big[\mu I\big(\hat{\mathbf{X}};\mathbf{V}\big) - I\big(\hat{\mathbf{Y}};\mathbf{V}\big)\big]. \tag{53}$$

We next define two subproblems that are used to lower bound the *enhanced optimization problem* $(\hat{P})$. The first subproblem $(\hat{P}_1)$ minimizes the first mutual information in the right-hand-side of (53) subject to the distortion constraint in $(\hat{P})$ and the second subproblem $(\hat{P}_2)$ maximizes the expression within the parenthesis in the right-hand-side of (53) subject to the Markov condition in $(\hat{P})$. In other words, $(\hat{P}_1)$ is defined as

$$\begin{aligned} (\hat{P}_1) \qquad \triangleq \qquad & \min_{\mathbf{U},\mathbf{V}} \ \mu I\big(\hat{\mathbf{X}};\mathbf{U},\mathbf{V}\big) \\ & \text{subject to} \ \ \mathbf{K}_{\hat{\mathbf{X}}|\mathbf{U},\mathbf{V}} \preccurlyeq \hat{\mathbf{D}}, \end{aligned}$$

and $(\hat{P}_2)$ is defined as

$$\begin{aligned} (\hat{P}_2) \qquad \triangleq \qquad & \max_{\mathbf{V}} \ \mu I\big(\hat{\mathbf{X}};\mathbf{V}\big) - I\big(\hat{\mathbf{Y}};\mathbf{V}\big) \\ & \text{subject to} \ \ \hat{\mathbf{X}} \leftrightarrow \hat{\mathbf{Y}} \leftrightarrow \mathbf{V}. \end{aligned}$$

It is clear from the decomposition in (53) and from the definitions of $(\hat{P})$, $(\hat{P}_1)$, and $(\hat{P}_2)$ that $(\hat{P}_1)$ and $(\hat{P}_2)$ lower bound $(\hat{P})$, i.e.,

$$v(\hat{P}) \geq v(\hat{P}_1) - v(\hat{P}_2). \tag{54}$$

We now give two lemmas about the optimal solutions to subproblems $(\hat{P}_1)$ and $(\hat{P}_2)$.

Lemma 6. *A Gaussian* $(\mathbf{U},\mathbf{V})$ *with the conditional covariance matrix*

$$\mathbf{K}_{\hat{\mathbf{X}}|\mathbf{U},\mathbf{V}} = \mathbf{K}_{\hat{\mathbf{X}}} - \tilde{\mathbf{B}}_1^* - \tilde{\mathbf{B}}_2^* = \hat{\mathbf{D}}$$

is optimal for the subproblem $(\hat{P}_1)$*, and the optimal value is*

$$v(\hat{P}_1) = \frac{\mu}{2}\log\frac{|\mathbf{K}_{\hat{\mathbf{X}}}|}{|\hat{\mathbf{D}}|}. \tag{55}$$

Proof. See Appendix F. □

Lemma 7. *A Gaussian* $\mathbf{V}$ *with the conditional covariance matrix*

$$\mathbf{K}_{\hat{\mathbf{Y}}|\mathbf{V}} = \mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*$$

is optimal for the subproblem $(\hat{P}_2)$*, and the optimal value is*

$$v(\hat{P}_2) = \frac{\mu}{2} \log \frac{\left|\mathbf{K}_{\hat{\mathbf{X}}}\right|}{\left|\mathbf{K}_{\hat{\mathbf{X}}} - \tilde{\mathbf{B}}_2^*\right|} - \frac{1}{2} \log \frac{\left|\mathbf{K}_{\hat{\mathbf{Y}}}\right|}{\left|\mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\right|}. \tag{56}$$

Proof. See Appendix G. □

Substituting (55) and (56) into (54), we obtain

$$\begin{aligned} v(\hat{P}) &\geq \frac{\mu}{2} \log \frac{\left|\mathbf{K}_{\hat{\mathbf{X}}} - \tilde{\mathbf{B}}_2^*\right|}{\left|\hat{\mathbf{D}}\right|} + \frac{1}{2} \log \frac{\left|\mathbf{K}_{\hat{\mathbf{Y}}}\right|}{\left|\mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\right|} \\ &= \frac{\mu}{2} \log \frac{\left|\mathbf{K}_{\tilde{\mathbf{X}}} - \tilde{\mathbf{B}}_2^*\right|}{\left|\tilde{\mathbf{D}}\right|} + \frac{1}{2} \log \frac{\left|\mathbf{K}_{\tilde{\mathbf{Y}}}\right|}{\left|\mathbf{K}_{\tilde{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\right|} && (57) \\ &= v\big(P_{G2}\big), && (58) \end{aligned}$$

where

(57) follows from (38), (45), (46), and (47), and

(58) follows from (26).

We conclude from (35), (52), and (58) that

$$v\,(P) \geq v\,(P_{G2})\,.$$

It now follows from this and (23) that

$$v\,(P) = v\,(P_{G1}) = v\,(P_{G2})\,, \tag{59}$$

which proves that a Gaussian $(\mathbf{U}, \mathbf{V})$ is optimal for the *main optimization problem* (P). This completes the proof of Theorem 3.

7 Converse Proof of Theorem 1

Liu and Viswanath gave a single-letter outer bound to the rate region in [4]. We shall use a similar outer bound that is reminiscent of the Berger-Tung outer bound [9, 10].

Lemma 8. *If the rate-distortion vector* $(R_1, R_2, \mathbf{D})$ *is achievable, then there exist random vectors* $\mathbf{U}$ *and* $\mathbf{V}$ *such that*

$$\begin{aligned} &R_1 \geq I(\mathbf{X}; \mathbf{U}|\mathbf{V}), \\ &R_2 \geq I(\mathbf{Y}; \mathbf{V}), \\ &\mathbf{D} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}}, \quad \textit{and} \\ &\mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{V}. \end{aligned}$$

The proof of the lemma is similar to [7, Lemma 2] and is omitted. We are now ready to prove the converse of Theorem 1. If $(R_1, R_2, \mathbf{D})$ is achievable, then

$$\begin{aligned} \mu R_1 + R_2 &\geq v\,(P) && (60) \\ &= \begin{cases} v\,(P_{pt-pt}) & \text{if } 0 \leq \mu \leq 1 \\ v\,(P_{G1}) & \text{if } \mu > 1 \end{cases} && (61) \\ &= \mathcal{R}^*(\mathbf{D}, \mu), && (62) \end{aligned}$$

where

(60) follows from Lemma 8, and

(61) follows from (22) and (59).

And if $(R_1, R_2, \mathbf{D}) \in \overline{\mathcal{RD}}$, then (62) again holds because $\mathcal{R}^*(\mathbf{D}, \mu)$ is continuous in $\mathbf{D}$. So, (62) is a lower bound for any (R_1, R_2) in the rate region $\mathcal{R}(\mathbf{D})$. Hence,

$$\begin{aligned}\mathcal{R}(\mathbf{D}, \mu) &= \inf_{(R_1,R_2)\in\mathcal{R}(\mathbf{D})} \mu R_1 + R_2 \\ &\geq \mathcal{R}^*(\mathbf{D}, \mu).\end{aligned}$$

This completes the converse proof of Theorem 1.

Remark 3: It follows from Theorem 1 and Lemma 1 that one can add the constraints

$$\begin{aligned}&\mathbf{U} \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{V} \quad \text{and} \\ &(\mathbf{U}, \mathbf{V}, \mathbf{X}, \mathbf{Y}) \text{ are jointly Gaussian}\end{aligned}$$

to the optimization problem

$$\begin{aligned}(P) \qquad \triangleq \qquad &\min_{\mathbf{U},\mathbf{V}} \; \mu I(\mathbf{X}; \mathbf{U}|\mathbf{V}) + I(\mathbf{Y}; \mathbf{V}) \\ &\text{subject to } \; \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} \preccurlyeq \mathbf{D} \quad \text{and} \\ &\qquad\qquad\quad \mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{V},\end{aligned}$$

without changing its optimal value.

8 Solution for the General Case

In this section, we lift the assumptions on $\mathbf{K_X}, \mathbf{K_Y}$, and $\mathbf{D}$ and allow them to be any positive semidefinite matrices. We shall show that the Gaussian achievable scheme is optimal for this general problem. For this section, we denote the rate region of the problem by $\mathcal{R}(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D})$. Note that $\mathbf{K_X}$ and $\mathbf{K_Y}$ completely specify the joint distribution of $\mathbf{X}$ and $\mathbf{Y}$ because we continue to assume that $\mathbf{X} = \mathbf{Y} + \mathbf{N}$. Similarly, $\mathcal{R}_G(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D})$ is used to denote the rate region achieved by the Gaussian achievable scheme. We use $\mathcal{R}(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu)$ and $\mathcal{R}_G(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu)$ to denote the two minimum weighted sum-rates. Likewise, we denote the set $\mathcal{S}$ defined in Section 3.2 by $\mathcal{S}(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D})$. We use similar notation later in the section. We start with the following extension.

Theorem 4. *If $\mathbf{K_X}$ and $\mathbf{D}$ are positive definite, and $\mathbf{K_Y}$ is positive semidefinite, then*

$$\mathcal{R}(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu) = \mathcal{R}_G(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu).$$

Proof. It suffices to prove that

$$\mathcal{R}(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu) \geq \mathcal{R}_G(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu).$$

If $\mathbf{K_Y}$ is positive definite (hence nonsingular), then the result follows from Theorem 1. We therefore assume that $\mathbf{K_Y}$ is singular and has a rank $p < m$. The eigen decomposition of $\mathbf{K_Y}$ is

$$\mathbf{K_Y} = \mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^T,$$

where $\mathbf{Q}$ is an orthogonal matrix and

$$\mathbf{\Sigma} = \text{Diag}(\alpha_1, \ldots, \alpha_p, 0, \ldots, 0).$$

Let us partition $\mathbf{Q}$ as

$$\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2],$$

where $\mathbf{Q}_1$ is an $m \times p$ matrix. Let us define

$$\mathbf{Q}^T\mathbf{K_N}\mathbf{Q} \triangleq \begin{pmatrix} \mathbf{E} & \mathbf{F}^T \\ \mathbf{F} & \mathbf{G} \end{pmatrix},$$

where $\mathbf{E}$, $\mathbf{F}$, and $\mathbf{G}$ are submatrices of dimensions $p\times p$, $(m-p)\times p$, and $(m-p)\times(m-p)$, respectively. Since $\mathbf{Q}_2^T\mathbf{K_Y}\mathbf{Q_2} = \mathbf{0}$ and $\mathbf{X} = \mathbf{Y} + \mathbf{N}$, we have that

$$\mathbf{G} = \mathbf{Q}_2^T\mathbf{K_N}\mathbf{Q_2} = \mathbf{Q}_2^T\mathbf{K_X}\mathbf{Q_2} \succ \mathbf{0},$$

i.e., $\mathbf{G}$ is positive definite. Using this, we define

$$\mathbf{A} \triangleq \begin{pmatrix} \mathbf{I}_p & -\mathbf{F}^T\mathbf{G}^{-1} \\ \mathbf{0} & \mathbf{I}_{m-p} \end{pmatrix} \mathbf{Q}^T.$$

$\mathbf{A}$ defines a transformed problem in which the transformed sources are

$$\begin{aligned} \bar{\mathbf{X}} &\triangleq \mathbf{A}\mathbf{X} \text{ and} \\ \bar{\mathbf{Y}} &\triangleq \mathbf{A}\mathbf{Y}, \end{aligned}$$

which satisfy

$$\bar{\mathbf{X}} = \bar{\mathbf{Y}} + \bar{\mathbf{N}},$$

where $\bar{\mathbf{N}} \triangleq \mathbf{A}\mathbf{N}$, and the transformed distortion matrix is

$$\bar{\mathbf{D}} \triangleq \mathbf{A}\mathbf{D}\mathbf{A}^T.$$

The covariance matrix of the transformed source $\bar{\mathbf{Y}}$ is

$$\mathbf{K}_{\bar{\mathbf{Y}}} = \mathbf{A}\mathbf{K_Y}\mathbf{A}^T = \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where

$$\mathbf{\Sigma}_1 \triangleq \text{Diag}(\alpha_1, \ldots, \alpha_p),$$

and the covariance matrix of $\bar{\mathbf{N}}$ is

$$\begin{aligned} \mathbf{K}_{\bar{\mathbf{N}}} &= \mathbf{A}\mathbf{K_N}\mathbf{A}^T \\ &= \begin{pmatrix} \mathbf{I}_p & -\mathbf{F}^T\mathbf{G}^{-1} \\ \mathbf{0} & \mathbf{I}_{m-p} \end{pmatrix} \begin{pmatrix} \mathbf{E} & \mathbf{F}^T \\ \mathbf{F} & \mathbf{G} \end{pmatrix} \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ -\mathbf{G}^{-1}\mathbf{F} & \mathbf{I}_{m-p} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{E} - \mathbf{F}^T\mathbf{G}^{-1}\mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{pmatrix}. \end{aligned}$$

Using these, the covariance matrix of the transformed source $\bar{\mathbf{X}}$ can be expressed as

$$\begin{aligned} \mathbf{K}_{\bar{\mathbf{X}}} &= \mathbf{K}_{\bar{\mathbf{Y}}} + \mathbf{K}_{\bar{\mathbf{N}}} \\ &= \begin{pmatrix} \mathbf{\Sigma}_1 + \mathbf{E} - \mathbf{F}^T\mathbf{G}^{-1}\mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{pmatrix}. \end{aligned}$$

Since $\mathbf{A}$ is invertible, the above transformation is information lossless, and hence the transformed problem is equivalent to the original problem. Therefore,

$$\begin{aligned} \mathcal{R}\left(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu\right) &= \mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right) \text{ and} \\ \mathcal{R}_G\left(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu\right) &= \mathcal{R}_G\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right). \end{aligned}$$

So, it is sufficient to prove that

$$\mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right) \geq \mathcal{R}_G\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right).$$

Let us define the following matrices

$$\begin{aligned} \mathbf{K}_{\bar{\mathbf{N}}_1^{(n)}} &\triangleq \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{n}\mathbf{G} \end{pmatrix} \text{ and} \\ \mathbf{K}_{\bar{\mathbf{N}}_2^{(n)}} &\triangleq \begin{pmatrix} \mathbf{E} - \mathbf{F}^T\mathbf{G}^{-1}\mathbf{F} & \mathbf{0} \\ \mathbf{0} & \left(1 - \frac{1}{n}\right)\mathbf{G} \end{pmatrix}, \end{aligned}$$

where n is a positive integer. It is clear that these matrices are positive semidefinite and they satisfy

$$\mathbf{K}_{\bar{\mathbf{N}}} = \mathbf{K}_{\bar{\mathbf{N}}_1^{(n)}} + \mathbf{K}_{\bar{\mathbf{N}}_2^{(n)}}.$$

Let $\bar{\mathbf{N}}_1^{(n)}$ and $\bar{\mathbf{N}}_2^{(n)}$ be zero-mean vector Gaussian sources with covariance matrices $\mathbf{K}_{\bar{\mathbf{N}}_1^{(n)}}$ and $\mathbf{K}_{\bar{\mathbf{N}}_2^{(n)}}$, respectively. In addition, suppose they are independent of each other and all other vector Gaussian sources. We can then write

$$\bar{\mathbf{X}} = \bar{\mathbf{Y}} + \bar{\mathbf{N}}_1^{(n)} + \bar{\mathbf{N}}_2^{(n)}.$$

Let us consider a new problem in which encoder 1 has access to $\bar{\mathbf{X}}$, encoder 2 has access to $\left(\bar{\mathbf{Y}}, \bar{\mathbf{N}}_1^{(n)}\right)$, and the distortion constraint on $\bar{\mathbf{X}}$ is $\bar{\mathbf{D}}$. This problem is clearly a relaxation to the original problem because encoder 2 has access to more information about $\bar{\mathbf{X}}$ than the original problem. In other words, any feasible scheme for the original problem is also feasible for this new problem. Now since there is no distortion constraint on $\bar{\mathbf{Y}}$ and the sufficient statistic of $\bar{\mathbf{X}}$ in $\left(\bar{\mathbf{Y}}, \bar{\mathbf{N}}_1^{(n)}\right)$ is $\bar{\mathbf{Y}} + \bar{\mathbf{N}}_1^{(n)}$, this new problem is equivalent to the problem in which encoder 2, instead of $\left(\bar{\mathbf{Y}}, \bar{\mathbf{N}}_1^{(n)}\right)$, has access to the sum $\bar{\mathbf{Y}} + \bar{\mathbf{N}}_1^{(n)}$. Let us denote this sum by $\bar{\mathbf{Y}}^{(n)}$, i.e.,

$$\bar{\mathbf{Y}}^{(n)} \triangleq \bar{\mathbf{Y}} + \bar{\mathbf{N}}_1^{(n)},$$

which has a positive definite covariance matrix

$$\mathbf{K}_{\bar{\mathbf{Y}}^{(n)}} = \mathbf{K}_{\bar{\mathbf{Y}}} + \mathbf{K}_{\bar{\mathbf{N}}_1^{(n)}} = \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \frac{1}{n}\mathbf{G} \end{pmatrix}.$$

It follows that

$$\mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}, \bar{\mathbf{D}}, \mu\right) \le \mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right).$$

Since this is true for all n and $\mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}, \bar{\mathbf{D}}, \mu\right)$ is nondecreasing in n, we obtain

$$\lim_{n\to\infty} \mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}, \bar{\mathbf{D}}, \mu\right) \le \mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right). \tag{63}$$

Since $\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}$, and $\bar{\mathbf{D}}$ are positive definite, the conclusion of Theorem 1 holds for this sequence of relaxed problems, i.e., for each n

$$\mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}, \bar{\mathbf{D}}, \mu\right) = \mathcal{R}_G\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}, \bar{\mathbf{D}}, \mu\right).$$

This and (63) together imply that

$$\lim_{n\to\infty} \mathcal{R}_G\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}, \bar{\mathbf{D}}, \mu\right) \le \mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right). \tag{64}$$

Now for each n, there exists $\left(\mathbf{U}^{(n)}, \mathbf{V}^{(n)}\right)$ in $\mathcal{S}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}, \bar{\mathbf{D}}\right)$ such that

$$\mathcal{R}_G\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}, \bar{\mathbf{D}}, \mu\right) = \mu I\left(\bar{\mathbf{X}}; \mathbf{U}^{(n)} | \mathbf{V}^{(n)}\right) + I\left(\bar{\mathbf{Y}}^{(n)}; \mathbf{V}^{(n)}\right). \tag{65}$$

Since $\bar{\mathbf{X}}, \bar{\mathbf{Y}}^{(n)}, \mathbf{U}^{(n)}$, and $\mathbf{V}^{(n)}$ are jointly Gaussian, we can without loss of generality parameterize them by positive semidefinite matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ as in the definition (P_{G2}). These matrices lie in a compact set because they satisfy the KKT conditions that are continuous, and they are bounded as $\mathbf{B}_1 + \mathbf{B}_2 \prec \mathbf{K}_{\bar{\mathbf{X}}}$. Therefore, there exists a subsequence of $\mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}$ along which $\left(\mathbf{U}^{(n)}, \mathbf{V}^{(n)}\right)$ converges to $(\mathbf{U}, \mathbf{V})$ in $\mathcal{S}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}\right)$. Since the right-hand-side of (65) is continuous in $\left(\bar{\mathbf{Y}}^{(n)}, \mathbf{U}^{(n)}, \mathbf{V}^{(n)}\right)$, this implies

$$\begin{aligned}\lim_{n\to\infty} \mathcal{R}_G\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}^{(n)}}, \bar{\mathbf{D}}, \mu\right) &= \mu I\left(\bar{\mathbf{X}}; \mathbf{U} | \mathbf{V}\right) + I\left(\bar{\mathbf{Y}}; \mathbf{V}\right) \\ &\ge \mathcal{R}_G\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right).\end{aligned} \tag{66}$$

It now follows from (64) and (66) that

$$\mathcal{R}\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right) \ge \mathcal{R}_G\left(\mathbf{K}_{\bar{\mathbf{X}}}, \mathbf{K}_{\bar{\mathbf{Y}}}, \bar{\mathbf{D}}, \mu\right).$$

This proves Theorem 4. $\square$

We next use Theorem 4 to prove our result for the most general case of the problem.

Theorem 5. *For any positive semidefinite* $\mathbf{K_X}, \mathbf{K_Y}$*, and* $\mathbf{D}$*, we have*

$$\mathcal{R}\left(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu\right) = \mathcal{R}_G\left(\mathbf{K_X}, \mathbf{K_Y}, \mathbf{D}, \mu\right).$$

Proof. Let us suppose that the rank of $\mathbf{K_X}$ is $p \leq m$. Since $\mathbf{K_X}$ is positive semidefinite, its eigen decomposition is

$$\mathbf{K_X} = \mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^T,$$

where $\mathbf{Q}$ is an orthogonal matrix and

$$\boldsymbol{\Sigma} = \text{Diag}(\alpha_1, \ldots, \alpha_p, 0, \ldots, 0).$$

Let us partition $\mathbf{Q}$ as

$$\mathbf{Q} \triangleq [\mathbf{Q}_1, \mathbf{Q}_2],$$

where $\mathbf{Q}_1$ is an $m \times p$ matrix. Since $\mathbf{Q}_2^T\mathbf{K_X}\mathbf{Q_2} = \mathbf{0}$ and $\mathbf{X} = \mathbf{Y} + \mathbf{N}$, we have

$$\mathbf{Q}_2^T\mathbf{K_Y}\mathbf{Q_2} = \mathbf{Q}_2^T\mathbf{K_N}\mathbf{Q_2} = \mathbf{0},$$

which implies that

$$\mathbf{Q}^T\mathbf{K_Y}\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1^T\mathbf{K_Y}\mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{and}$$
$$\mathbf{Q}^T\mathbf{K_N}\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1^T\mathbf{K_N}\mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Let us define

$$\mathbf{Q}^T\mathbf{D}\mathbf{Q} \triangleq \begin{pmatrix} \mathbf{E} & \mathbf{F}^T \\ \mathbf{F} & \mathbf{G} \end{pmatrix},$$

where $\mathbf{E}$, $\mathbf{F}$, and $\mathbf{G}$ are submatrices of dimensions $p \times p$, $(m-p) \times p$, and $(m-p) \times (m-p)$, respectively. We need the following lemma.

Lemma 9. *[20, Appendix A.5.5, p. 651]* $\mathbf{Q}^T\mathbf{D}\mathbf{Q} \succcurlyeq \mathbf{0}$ *if and only if*

$$\begin{aligned} \mathbf{G} &\succcurlyeq \mathbf{0}, \\ \mathbf{E} - \mathbf{F}^T\mathbf{G}^+\mathbf{F} &\succcurlyeq \mathbf{0}, \ \textit{and} \\ (\mathbf{I}_{m-p} - \mathbf{G}\mathbf{G}^+)\mathbf{F} &= \mathbf{0}, \end{aligned}$$

where $\mathbf{G}^+$ *is the pseudo-inverse or Moore-Penrose inverse of* $\mathbf{G}$ *[20, Appendix A.5.4, p. 649].*

Let

$$\mathbf{T} \triangleq \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{I}_p & -\mathbf{F}^T\mathbf{G}^+ \\ \mathbf{0} & \mathbf{I}_{m-p} \end{pmatrix} \mathbf{Q}^T,$$

where $\mathbf{T}_1$ is a $p \times m$ matrix. Using this, we obtain a transformed problem in which the transformed sources are

$$\bar{\mathbf{X}} \triangleq \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{T}_1\mathbf{X} \\ \mathbf{T}_2\mathbf{X} \end{pmatrix} = \mathbf{T}\mathbf{X} \quad \text{and}$$
$$\bar{\mathbf{Y}} \triangleq \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{T}_1\mathbf{Y} \\ \mathbf{T}_2\mathbf{Y} \end{pmatrix} = \mathbf{T}\mathbf{Y}.$$

Using Lemma 9, we obtain the transformed distortion matrix

$$\begin{aligned}
\bar{\mathbf{D}} &\triangleq \mathbf{T}\mathbf{D}\mathbf{T}^T \\
&= \begin{pmatrix} \mathbf{I}_p & -\mathbf{F}^T\mathbf{G}^+ \\ \mathbf{0} & \mathbf{I}_{m-p} \end{pmatrix} \mathbf{Q}^T\mathbf{D}\mathbf{Q} \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ -\mathbf{G}^+\mathbf{F} & \mathbf{I}_{m-p} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I}_p & -\mathbf{F}^T\mathbf{G}^+ \\ \mathbf{0} & \mathbf{I}_{m-p} \end{pmatrix} \begin{pmatrix} \mathbf{E} & \mathbf{F}^T \\ \mathbf{F} & \mathbf{G} \end{pmatrix} \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ -\mathbf{G}^+\mathbf{F} & \mathbf{I}_{m-p} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{E} - \mathbf{F}^T\mathbf{G}^+\mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix},
\end{aligned} \tag{67}$$

where

$$\begin{aligned}
\mathbf{D}_1 &\triangleq \mathbf{E} - \mathbf{F}^T\mathbf{G}^+\mathbf{F} \quad \text{and} \\
\mathbf{D}_2 &\triangleq \mathbf{G}.
\end{aligned}$$

The covariance matrix of the transformed source $\bar{\mathbf{X}}$ is

$$\begin{aligned}
\mathbf{K}_{\bar{\mathbf{X}}} &= \mathbf{T}\mathbf{K}_{\mathbf{X}}\mathbf{T}^T \\
&= \begin{pmatrix} \mathbf{I}_p & -\mathbf{F}^T\mathbf{G}^+ \\ \mathbf{0} & \mathbf{I}_{m-p} \end{pmatrix} \mathbf{Q}^T\mathbf{K}_{\mathbf{X}}\mathbf{Q} \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ -\mathbf{G}^+\mathbf{F} & \mathbf{I}_{m-p} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I}_p & -\mathbf{F}^T\mathbf{G}^+ \\ \mathbf{0} & \mathbf{I}_{m-p} \end{pmatrix} \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ -\mathbf{G}^+\mathbf{F} & \mathbf{I}_{m-p} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},
\end{aligned}$$

where

$$\mathbf{\Sigma}_1 \triangleq \text{Diag}(\alpha_1, \ldots, \alpha_p),$$

and the covariance matrix of the transformed source $\bar{\mathbf{Y}}$ is

$$\begin{aligned}
\mathbf{K}_{\bar{\mathbf{Y}}} &= \mathbf{T}\mathbf{K}_{\mathbf{Y}}\mathbf{T}^T \\
&= \begin{pmatrix} \mathbf{I}_p & -\mathbf{F}^T\mathbf{G}^+ \\ \mathbf{0} & \mathbf{I}_{m-p} \end{pmatrix} \mathbf{Q}^T\mathbf{K}_{\mathbf{Y}}\mathbf{Q} \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ -\mathbf{G}^+\mathbf{F} & \mathbf{I}_{m-p} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I}_p & -\mathbf{F}^T\mathbf{G}^+ \\ \mathbf{0} & \mathbf{I}_{m-p} \end{pmatrix} \begin{pmatrix} \mathbf{Q}_1^T\mathbf{K}_{\mathbf{Y}}\mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{I}_p & \mathbf{0} \\ -\mathbf{G}^+\mathbf{F} & \mathbf{I}_{m-p} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{Q}_1^T\mathbf{K}_{\mathbf{Y}}\mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.
\end{aligned}$$

It follows that $\mathbf{X}_2$ and $\mathbf{Y}_2$ are deterministic, i.e.,

$$\mathbf{X}_2 = \mathbf{Y}_2 = \mathbf{0}.$$

Since $\mathbf{T}$ is invertible, the distortion constraint is equivalent to

$$\begin{aligned}
\mathbf{T}\mathbf{D}\mathbf{T}^T &\succcurlyeq \frac{1}{n}\sum_{i=1}^{n} E\left[\left(\bar{\mathbf{X}}^n(i) - \hat{\bar{\mathbf{X}}}^n(i)\right)\left(\bar{\mathbf{X}}^n(i) - \hat{\bar{\mathbf{X}}}^n(i)\right)^T\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left[\begin{pmatrix} \mathbf{X}_1^n(i) - \hat{\mathbf{X}}_1^n(i) \\ \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{X}_1^n(i) - \hat{\mathbf{X}}_1^n(i) \\ \mathbf{0} \end{pmatrix}^T\right] \\
&= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} E\left[\left(\mathbf{X}_1^n(i) - \hat{\mathbf{X}}_1^n(i)\right)\left(\mathbf{X}_1^n(i) - \hat{\mathbf{X}}_1^n(i)\right)^T\right] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.
\end{aligned} \tag{68}$$

Since $\mathbf{D}_1$ and $\mathbf{D}_2$ are positive semidefinite from Lemma 9, (67) and (68) imply that the distortion constraint is equivalent to

$$\mathbf{D}_1 \succcurlyeq \frac{1}{n}\sum_{i=1}^{n} E\left[\left(\mathbf{X}_1^n(i) - \hat{\mathbf{X}}_1^n(i)\right)\left(\mathbf{X}_1^n(i) - \hat{\mathbf{X}}_1^n(i)\right)^T\right].$$

Since $\mathbf{T}$ is invertible, the above transformation is information lossless, and hence the transformed problem is equivalent to the original problem. Moreover, the transformed problem is effectively p-dimensional with the sources $\mathbf{X}_1$ and $\mathbf{Y}_1$, and the distortion matrix $\mathbf{D}_1$ such that

$$\mathbf{K}_{\mathbf{X}_1} = \mathbf{\Sigma}_1 \succ \mathbf{0} \quad \text{and}$$
$$\mathbf{X}_1 = \mathbf{Y}_1 + \mathbf{N}_1,$$

where $\mathbf{N}_1 \triangleq \mathbf{T}_1\mathbf{N}$. We therefore have that

$$\mathcal{R}\left(\mathbf{K}_{\mathbf{X}}, \mathbf{K}_{\mathbf{Y}}, \mathbf{D}, \mu\right) = \mathcal{R}\left(\mathbf{K}_{\mathbf{X}_1}, \mathbf{K}_{\mathbf{Y}_1}, \mathbf{D}_1, \mu\right) \quad \text{and} \tag{69}$$
$$\mathcal{R}_G\left(\mathbf{K}_{\mathbf{X}}, \mathbf{K}_{\mathbf{Y}}, \mathbf{D}, \mu\right) = \mathcal{R}_G\left(\mathbf{K}_{\mathbf{X}_1}, \mathbf{K}_{\mathbf{Y}_1}, \mathbf{D}_1, \mu\right). \tag{70}$$

Since $\mathbf{K}_{\mathbf{X}_1}$ is positive definite, if $\mathbf{D}_1$ is singular, then the right-hand side of (69) and (70) are both infinite, so the conclusion trivially holds. Otherwise, we have that $\mathbf{K}_{\mathbf{X}_1}$ and $\mathbf{D}_1$ are positive definite and $\mathbf{K}_{\mathbf{Y}_1}$ is positive semidefinite. In that case Theorem 4 implies that

$$\mathcal{R}\left(\mathbf{K}_{\mathbf{X}_1}, \mathbf{K}_{\mathbf{Y}_1}, \mathbf{D}_1, \mu\right) = \mathcal{R}_G\left(\mathbf{K}_{\mathbf{X}_1}, \mathbf{K}_{\mathbf{Y}_1}, \mathbf{D}_1, \mu\right).$$

This together with (69) and (70) establishes the desired equality

$$\mathcal{R}\left(\mathbf{K}_{\mathbf{X}}, \mathbf{K}_{\mathbf{Y}}, \mathbf{D}, \mu\right) = \mathcal{R}_G\left(\mathbf{K}_{\mathbf{X}}, \mathbf{K}_{\mathbf{Y}}, \mathbf{D}, \mu\right).$$

Theorem 5 is thus proved. □

Appendix A: Proof of Equality in Lemma 1

Suppose μ is in $[0, 1]$. Then for any $(\mathbf{U}, \mathbf{V})$ in $\mathcal{S}$, we have

$$\begin{aligned}
\mu I(\mathbf{X};\mathbf{U}|\mathbf{V}) + I(\mathbf{Y};\mathbf{V}) &= \mu I(\mathbf{X};\mathbf{U},\mathbf{V}) - \mu I(\mathbf{X},\mathbf{V}) + I(\mathbf{Y};\mathbf{V}) \\
&= \mu I(\mathbf{X};\mathbf{U}) + \mu I(\mathbf{X};\mathbf{V}|\mathbf{U}) + \mu[I(\mathbf{Y},\mathbf{V}) - I(\mathbf{X};\mathbf{V})] + (1-\mu)I(\mathbf{Y};\mathbf{V}) \\
&\geq \mu I(\mathbf{X};\mathbf{U}) \qquad (71) \\
&= \frac{\mu}{2}\log\frac{|\mathbf{K}_{\mathbf{X}}|}{|\mathbf{K}_{\mathbf{X}|\mathbf{U}}|}
\end{aligned}$$

where (71) follows because of the facts that

$$I(\mathbf{Y};\mathbf{V}) \geq 0$$

and

$$I(\mathbf{X};\mathbf{V}|\mathbf{U}) \geq 0,$$

and we have

$$I(\mathbf{Y},\mathbf{V}) - I(\mathbf{X};\mathbf{V}) \geq 0$$

because of the data processing inequality [21, Theorem 2.8.1] and the Markov chain $\mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{V}$. The inequality (71) is achieved by any $(\mathbf{U}, \mathbf{V})$ in $\mathcal{S}$ such that $\mathbf{V}$ is independent of $(\mathbf{X}, \mathbf{Y}, \mathbf{U})$, and the conditional covariance of $\mathbf{X}$ given $(\mathbf{U}, \mathbf{V})$ satisfies

$$\mathbf{0} \preccurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} = \mathbf{K}_{\mathbf{X}|\mathbf{U}} \preccurlyeq \mathbf{D}.$$

Since conditioning reduces covariance in a positive semidefinite sense, we have an additional constraint

$$\mathbf{K}_{\mathbf{X}|\mathbf{U}} \preccurlyeq \mathbf{K}_{\mathbf{X}}.$$

We therefore have the following

$$
\begin{aligned}
\mathcal{R}_G(\mathbf{D},\mu) &= \min_{(R_1,R_2)\in\mathcal{R}_G(\mathbf{D})} \mu R_1 + R_2 \\
&= \min_{(\mathbf{U},\mathbf{V})\in\mathcal{S}} \mu I(\mathbf{X};\mathbf{U}|\mathbf{V}) + I(\mathbf{Y};\mathbf{V}) \\
&= \begin{array}{cl} \min\limits_{\mathbf{K}_{\mathbf{X}|\mathbf{U}}} & \dfrac{\mu}{2}\log\dfrac{|\mathbf{K}_{\mathbf{X}}|}{|\mathbf{K}_{\mathbf{X}|\mathbf{U}}|} \\ \text{subject to} & \mathbf{K}_{\mathbf{X}} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U}} \succcurlyeq \mathbf{0} \quad \text{and} \\ & \mathbf{D} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U}} \end{array} \\
&= v\left(P_{pt-pt}\right).
\end{aligned}
$$

Suppose now that $\mu > 1$. Then any $(\mathbf{U},\mathbf{V})$ in $\mathcal{S}$ can be characterized by positive semidefinite conditional covariance matrices $\mathbf{K}_{\mathbf{Y}|\mathbf{V}}$ and $\mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}}$ such that

$$
\begin{aligned}
&\mathbf{K}_{\mathbf{Y}} \succcurlyeq \mathbf{K}_{\mathbf{Y}|\mathbf{V}} \succcurlyeq \mathbf{0}, \\
&\mathbf{K}_{\mathbf{Y}|\mathbf{V}} + \mathbf{K}_{\mathbf{N}} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} \succcurlyeq \mathbf{0}, \\
&\mathbf{D} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}},
\end{aligned}
$$

and

$$
\begin{aligned}
I(\mathbf{X};\mathbf{U}|\mathbf{V}) &= \frac{1}{2}\log\frac{|\mathbf{K}_{\mathbf{Y}|\mathbf{V}} + \mathbf{K}_{\mathbf{N}}|}{|\mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}}|}, \\
I(\mathbf{Y};\mathbf{V}) &= \frac{1}{2}\log\frac{|\mathbf{K}_{\mathbf{Y}}|}{|\mathbf{K}_{\mathbf{Y}|\mathbf{V}}|}.
\end{aligned}
$$

In this case, we have

$$
\begin{aligned}
\mathcal{R}_G(\mathbf{D},\mu) &= \min_{(R_1,R_2)\in\mathcal{R}_G(\mathbf{D})} \mu R_1 + R_2 \\
&= \min_{(\mathbf{U},\mathbf{V})\in\mathcal{S}} \mu I(\mathbf{X};\mathbf{U}|\mathbf{V}) + I(\mathbf{Y};\mathbf{V}) \\
&= \begin{array}{cl} \min\limits_{\mathbf{K}_{\mathbf{Y}|\mathbf{V}},\mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}}} & \dfrac{\mu}{2}\log\dfrac{|\mathbf{K}_{\mathbf{Y}|\mathbf{V}} + \mathbf{K}_{\mathbf{N}}|}{|\mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}}|} + \dfrac{1}{2}\log\dfrac{|\mathbf{K}_{\mathbf{Y}}|}{|\mathbf{K}_{\mathbf{Y}|\mathbf{V}}|} \\ \text{subject to} & \mathbf{K}_{\mathbf{Y}} \succcurlyeq \mathbf{K}_{\mathbf{Y}|\mathbf{V}} \succcurlyeq \mathbf{0}, \\ & \mathbf{K}_{\mathbf{Y}|\mathbf{V}} + \mathbf{K}_{\mathbf{N}} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} \succcurlyeq \mathbf{0}, \quad \text{and} \\ & \mathbf{D} \succcurlyeq \mathbf{K}_{\mathbf{X}|\mathbf{U},\mathbf{V}} \end{array} \\
&= v\left(P_{G1}\right).
\end{aligned}
$$

Appendix B: Proof of Lemma 2

We will be using several results and terms from Bertsekas *et al.* [22]. The book contains all of the background that these results need. The proof of the lemma is partially similar to that of Lemma 5 in [3]. Let us first introduce some notation used in the proof. We use vec$(\mathbf{A}_1,\mathbf{A}_2)$ to denote the column vector created by the concatenation of the columns of $m\times m$ matrices $\mathbf{A}_1$ and $\mathbf{A}_2$. If $\mathbf{a} = \text{vec}(\mathbf{A}_1,\mathbf{A}_2)$, then we use the notation mat$(\mathbf{a})$ to denote the inverse operation to get back the pair $(\mathbf{A}_1,\mathbf{A}_2)$, i.e.,

$$\text{mat}(\mathbf{a}) = (\mathbf{A}_1,\mathbf{A}_2).$$

The set of all column vectors created by the concatenation of the columns of $m\times m$ symmetric matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ is denoted by $\mathcal{A}$, i.e.,

$$\mathcal{A} \triangleq \{\text{vec}(\mathbf{A}_1,\mathbf{A}_2) : \mathbf{A}_i = \mathbf{A}_i^T \text{ for all } i\in\{1,2\}\}.$$

ri$(\mathcal{B})$ is used to denote the relative interior of the set $\mathcal{B}$. The sum of the two vector sets $\mathcal{V}_1$ and $\mathcal{V}_1$ is denoted by $\mathcal{V}_1+\mathcal{V}_2$ and is defined as

$$\mathcal{V}_1+\mathcal{V}_2 \triangleq \{\mathbf{v}_1+\mathbf{v}_2 : \mathbf{v}_i\in\mathcal{V}_i \text{ for all } i\in\{1,2\}\}.$$

We also need the following facts from linear algebra.

Lemma 10. *(a) If* $\mathbf{E}$ *is an* $m \times n$ *matrix and* $\mathbf{F}$ *is an* $n \times m$ *matrix, then* $\text{Tr}(\mathbf{EF}) = \text{Tr}(\mathbf{FE})$.

(b) If $\mathbf{E}$ *and* $\mathbf{F}$ *are positive semidefinite, then* $\mathbf{EF} = \mathbf{0}$ *if and only if* $\text{Tr}(\mathbf{EF}) = 0$.

Proof. Part (a) immediately follows from the definition of $\text{Tr}(\cdot)$ function. Part (b) can be proved using the eigen decompositions of $\mathbf{E}$ and $\mathbf{F}$. □

We can express the problem (P_{G2}) as

$$\begin{aligned} \min_{\mathbf{b}} \quad & h(\mathbf{b}) \\ \text{subject to} \quad & \mathbf{b} \in \mathcal{B}, \end{aligned}$$

where $\mathbf{b} \triangleq \text{vec}(\mathbf{B}_1, \mathbf{B}_2)$,

$$h(\mathbf{b}) \triangleq \frac{\mu}{2} \log \frac{|\mathbf{K_X} - \mathbf{B}_2|}{|\mathbf{K_X} - \mathbf{B_1} - \mathbf{B}_2|} + \frac{1}{2} \log \frac{|\mathbf{K_Y}|}{|\mathbf{K_Y} - \mathbf{B}_2|},$$

and the feasible set $\mathcal{B}$ is written as

$$\mathcal{B} \triangleq \mathcal{B}_1 \cap \mathcal{B}_2 \cap \mathcal{B}_{12},$$

where for $i \in \{1, 2\}$

$$\mathcal{B}_i \triangleq \{\text{vec}(\mathbf{B}_1, \mathbf{B}_2) : \mathbf{B}_i \succcurlyeq \mathbf{0}\} \cap \mathcal{A}$$

and

$$\mathcal{B}_{12} \triangleq \{\text{vec}(\mathbf{B}_1, \mathbf{B}_2) : \mathbf{B}_1 + \mathbf{B}_2 \succcurlyeq \mathbf{K_X} - \mathbf{D}\} \cap \mathcal{A}.$$

Since $h(\cdot)$ is continuously differentiable, it follows from [22, Proposition 4.7.1, p. 255] that any local minima $\mathbf{b}^*$ must satisfy

$$-\nabla h(\mathbf{b}^*) \in T_{\mathcal{B}}(\mathbf{b}^*)^*, \tag{72}$$

where $\nabla h(\mathbf{b}^*)$ is the gradient of $h(\cdot)$ at $\mathbf{b}^*$, and $T_{\mathcal{B}}(\mathbf{b}^*)^*$ is the polar cone of the tangent cone $T_{\mathcal{B}}(\mathbf{b}^*)$ of $\mathcal{B}$ at $\mathbf{b}^*$. Now since $\mathcal{B}_i$ for all $i \in \{1, 2\}$ and $\mathcal{B}_{12}$ are nonempty convex sets and $\text{ri}(\mathcal{B}_1) \cap \text{ri}(\mathcal{B}_2) \cap \text{ri}(\mathcal{B}_{12})$ is nonempty, it follows from [22, Problem 4.23, p. 267] and [22, Proposition 4.6.3, p. 254] that

$$T_{\mathcal{B}}(\mathbf{b}^*)^* = T_{\mathcal{B}_1}(\mathbf{b}^*)^* + T_{\mathcal{B}_2}(\mathbf{b}^*)^* + T_{\mathcal{B}_{12}}(\mathbf{b}^*)^*. \tag{73}$$

We next show that

$$-\nabla h(\mathbf{b}^*) \in T_{\mathcal{B}_1}(\mathbf{b}^*)^* \cap \mathcal{A} + T_{\mathcal{B}_2}(\mathbf{b}^*)^* \cap \mathcal{A} + T_{\mathcal{B}_{12}}(\mathbf{b}^*)^* \cap \mathcal{A}. \tag{74}$$

Note that $-\nabla h(\mathbf{b}^*)$ is a column concatenation of two $m \times m$ symmetric matrices. This together with (72) and (73) yields

$$-\nabla h(\mathbf{b}^*) = \mathbf{z}_1 + \mathbf{z}_2 + \mathbf{z}_{12} \in \mathcal{A}, \tag{75}$$

where for $i \in \{1, 2\}$

$$\begin{aligned} \mathbf{z}_i &\in T_{\mathcal{B}_i}(\mathbf{b}^*)^* \text{ and} \\ \mathbf{z}_{12} &\in T_{\mathcal{B}_{12}}(\mathbf{b}^*)^*. \end{aligned}$$

Let us now define

$$\begin{aligned} (\mathbf{K}_i, \mathbf{L}_i) &\triangleq \text{mat}(\mathbf{z}_i), \forall i \in \{1, 2\} \text{ and} \\ (\mathbf{K}_{12}, \mathbf{L}_{12}) &\triangleq \text{mat}(\mathbf{z}_{12}). \end{aligned}$$

Using this, we define

$$\begin{aligned} \bar{\mathbf{z}}_i &\triangleq \text{vec}\left(\frac{1}{2}\left(\mathbf{K}_i + \mathbf{K}_i^T\right), \frac{1}{2}\left(\mathbf{L}_i + \mathbf{L}_i^T\right)\right), \forall i \in \{1, 2\} \text{ and} \\ \bar{\mathbf{z}}_{12} &\triangleq \text{vec}\left(\frac{1}{2}\left(\mathbf{K}_{12} + \mathbf{K}_{12}^T\right), \frac{1}{2}\left(\mathbf{L}_{12} + \mathbf{L}_{12}^T\right)\right). \end{aligned}$$

Since $\mathcal{B}_1$ is a nonempty convex set, it follows from [22, Proposition 4.6.3, p. 254] that

$$\mathbf{z}_1^T(\mathbf{b}-\mathbf{b}^*) \leq 0, \quad \forall \mathbf{b} \in \mathcal{B}_1. \tag{76}$$

Consider any $\mathbf{b} \in \mathcal{B}_1$. Let

$$(\mathbf{E}_1, \mathbf{F}_1) \triangleq \text{mat}(\mathbf{b}-\mathbf{b}^*).$$

We now obtain

$$\begin{aligned}
\bar{\mathbf{z}}_1^T(\mathbf{b}-\mathbf{b}^*) &= \frac{1}{2}\text{Tr}\left(\left(\mathbf{K}_1+\mathbf{K}_1^T\right)\mathbf{E}_1\right) + \frac{1}{2}\text{Tr}\left(\left(\mathbf{L}_1+\mathbf{L}_1^T\right)\mathbf{F}_1\right) \\
&= \text{Tr}\left(\mathbf{K}_1\mathbf{E}_1\right) + \text{Tr}\left(\mathbf{L}_1\mathbf{F}_1\right) && (77) \\
&= \mathbf{z}_1^T(\mathbf{b}-\mathbf{b}^*) \\
&\leq 0, && (78)
\end{aligned}$$

where

(77) follows because $\mathbf{E}_1$ and $\mathbf{F}_1$ are symmetric, and

(78) follows from (76).

By definition, $\bar{\mathbf{z}}_1 \in \mathcal{A}$. This and (78) imply that

$$\bar{\mathbf{z}}_1 \in T_{\mathcal{B}_1}(\mathbf{b}^*)^* \cap \mathcal{A}. \tag{79}$$

We can similarly show that

$$\begin{aligned}
\bar{\mathbf{z}}_2 &\in T_{\mathcal{B}_2}(\mathbf{b}^*)^* \cap \mathcal{A} \text{ and} && (80) \\
\bar{\mathbf{z}}_{12} &\in T_{\mathcal{B}_{12}}(\mathbf{b}^*)^* \cap \mathcal{A}. && (81)
\end{aligned}$$

Now

$$\begin{aligned}
&\bar{\mathbf{z}}_1 + \bar{\mathbf{z}}_2 + \bar{\mathbf{z}}_{12} \\
&= \text{vec}\left(\frac{1}{2}\left(\mathbf{K}_1+\mathbf{K}_2+\mathbf{K}_{12}+\mathbf{K}_1^T+\mathbf{K}_2^T+\mathbf{K}_{12}^T\right), \frac{1}{2}\left(\mathbf{L}_1+\mathbf{L}_2+\mathbf{L}_{12}+\mathbf{L}_1^T+\mathbf{L}_2^T+\mathbf{L}_{12}^T\right)\right) \\
&= \text{vec}\left(\left(\mathbf{K}_1+\mathbf{K}_2+\mathbf{K}_{12}\right), \left(\mathbf{L}_1+\mathbf{L}_2+\mathbf{L}_{12}\right)\right) && (82) \\
&= \mathbf{z}_1+\mathbf{z}_2+\mathbf{z}_{12} \\
&= -\nabla h(\mathbf{b}^*), && (83)
\end{aligned}$$

where

(82) follows because $\mathbf{K}_1+\mathbf{K}_2+\mathbf{K}_{12}$ and $\mathbf{L}_1+\mathbf{L}_2+\mathbf{L}_{12}$ are symmetric from (75), and

(83) follows from the equality in (75).

This together with (79) – (81) implies (74).

We now proceed to characterize the right-hand side of (74). Consider any $\mathbf{z} \in T_{\mathcal{B}_1}(\mathbf{b}^*)^* \cap \mathcal{A}$. It again follows from [22, Proposition 4.6.3, p. 254] that

$$\mathbf{z}^T(\mathbf{b}-\mathbf{b}^*) \leq 0, \quad \forall \mathbf{b} \in \mathcal{B}_1. \tag{84}$$

Let us define

$$\begin{aligned}
(\mathbf{M}_1, \mathbf{M}_2) &\triangleq \text{mat}(\mathbf{z}), \\
(\mathbf{B}_1, \mathbf{B}_2) &\triangleq \text{mat}(\mathbf{b}), \text{ and} \\
(\mathbf{B}_1^*, \mathbf{B}_2^*) &\triangleq \text{mat}(\mathbf{b}^*).
\end{aligned}$$

Then (84) can be re-written as

$$\sum_{i=1}^{2} \mathrm{Tr}(\mathbf{M}_i(\mathbf{B}_i - \mathbf{B}_i^*)) \leq 0, \quad \forall\, \mathrm{vec}(\mathbf{B}_1, \mathbf{B}_2) \in \mathcal{B}_1. \tag{85}$$

We first show that $\mathbf{M}_2 = \mathbf{0}$. Let us pick $(\mathbf{B}_1, \mathbf{B}_2) = (\mathbf{B}_1^*, \mathbf{B}_2^* + \mathbf{M}_2)$. This means that

$$\mathrm{Tr}(\mathbf{M}_2\mathbf{M}_2) \leq 0,$$

which implies that $\mathbf{M}_2 = \mathbf{0}$ because $\mathbf{M}_2$ is symmetric. We next prove that $\mathbf{M}_1$ is negative semidefinite. Suppose there exists $\mathbf{w} \neq \mathbf{0}$ such that $\mathbf{w}^T\mathbf{M}_1\mathbf{w} > 0$. We then have

$$0 < \mathbf{w}^T\mathbf{M}_1\mathbf{w} = \mathrm{Tr}(\mathbf{w}^T\mathbf{M}_1\mathbf{w}) = \mathrm{Tr}(\mathbf{M}_1\mathbf{w}\mathbf{w}^T),$$

where the last equality follows from Lemma 10(a). But this contradicts (85) because $\mathrm{vec}(\mathbf{B}_1^* + \mathbf{w}\mathbf{w}^T, \mathbf{B}_2^*) \in \mathcal{B}_1$, and hence $\mathbf{M}_1 \preccurlyeq \mathbf{0}$. We finally show that $\mathbf{M}_1\mathbf{B}_1^* = \mathbf{0}$. Let $(\mathbf{B}_1, \mathbf{B}_2) = (\alpha\mathbf{B}_1^*, \mathbf{B}_2^*)$, where $\alpha > 1$. Then (85) implies that

$$\mathrm{Tr}(\mathbf{M}_1\mathbf{B}_1^*) \leq 0.$$

Likewise, on picking $0 < \alpha < 1$, we obtain

$$\mathrm{Tr}(\mathbf{M}_1\mathbf{B}_1^*) \geq 0.$$

Both together establish

$$\mathrm{Tr}(\mathbf{M}_1\mathbf{B}_1^*) = 0,$$

which together with Lemma 10(b) implies that

$$\mathbf{M}_1\mathbf{B}_1^* = \mathbf{0}$$

because $-\mathbf{M}_1$ and $\mathbf{B}_1^*$ are positive semidefinite. We therefore have that

$$T_{\mathcal{B}_1}(\mathbf{b}^*)^* \cap \mathcal{A} \subseteq \{\mathrm{vec}(\mathbf{M}_1, \mathbf{0}) : \mathbf{M}_1 \preccurlyeq \mathbf{0} \text{ and } \mathbf{M}_1\mathbf{B}_1^* = \mathbf{0}\}. \tag{86}$$

Similarly, we can show that

$$T_{\mathcal{B}_2}(\mathbf{b}^*)^* \cap \mathcal{A} \subseteq \{\mathrm{vec}(\mathbf{0}, \mathbf{M}_2) : \mathbf{M}_2 \preccurlyeq \mathbf{0} \text{ and } \mathbf{M}_2\mathbf{B}_2^* = \mathbf{0}\}. \tag{87}$$

Consider any $\mathbf{z} \in T_{\mathcal{B}_{12}}(\mathbf{b}^*)^* \cap \mathcal{A}$. As before, we obtain

$$\sum_{i=1}^{2} \mathrm{Tr}(\mathbf{\Lambda}_i(\mathbf{B}_i - \mathbf{B}_i^*)) \leq 0, \quad \forall\, \mathrm{vec}(\mathbf{B}_1, \mathbf{B}_2) \in \mathcal{B}_{12}, \tag{88}$$

where

$$(\mathbf{\Lambda}_1, \mathbf{\Lambda}_2) \triangleq \mathrm{mat}(\mathbf{z}).$$

On picking $(\mathbf{B}_1, \mathbf{B}_2) = (\mathbf{B}_1^* + \mathbf{\Lambda}_1, \mathbf{B}_2^* - \mathbf{\Lambda}_1)$, (88) yields

$$\mathrm{Tr}(\mathbf{\Lambda}_1\mathbf{\Lambda}_1) - \mathrm{Tr}(\mathbf{\Lambda}_2\mathbf{\Lambda}_1) \leq 0.$$

Similarly, picking $(\mathbf{B}_1, \mathbf{B}_2) = (\mathbf{B}_1^* - \mathbf{\Lambda}_2, \mathbf{B}_2^* + \mathbf{\Lambda}_2)$ gives

$$\mathrm{Tr}(\mathbf{\Lambda}_2\mathbf{\Lambda}_2) - \mathrm{Tr}(\mathbf{\Lambda}_1\mathbf{\Lambda}_2) \leq 0.$$

Both together imply that

$$\mathrm{Tr}((\mathbf{\Lambda}_1 - \mathbf{\Lambda}_2)(\mathbf{\Lambda}_1 - \mathbf{\Lambda}_2)) \leq 0,$$

and therefore

$$\mathbf{\Lambda}_1 - \mathbf{\Lambda}_2 = \mathbf{0},$$

because $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are symmetric. Let us denote $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ by $\mathbf{\Lambda}$. As before, we can show that $\mathbf{\Lambda} \preccurlyeq \mathbf{0}$. We next prove that

$$\mathrm{Tr}(\mathbf{\Lambda}(\mathbf{B}_1^* + \mathbf{B}_2^* - \mathbf{K}_{\mathbf{X}} + \mathbf{D})) = 0.$$

Observe that $(\mathbf{B}_1, \mathbf{B}_2) = \big(\alpha(\mathbf{B}_1^* + \mathbf{B}_2^* - \mathbf{K}_\mathbf{X} + \mathbf{D}) + \mathbf{K}_\mathbf{X} - \mathbf{D} - \mathbf{B}_2^*, \mathbf{B}_2^*\big)$, where $\alpha > 0$, is a valid choice of $(\mathbf{B}_1, \mathbf{B}_2)$ in (88). For $\alpha > 1$, this implies

$$\mathrm{Tr}(\mathbf{\Lambda}(\mathbf{B}_1^* + \mathbf{B}_2^* - \mathbf{K}_\mathbf{X} + \mathbf{D})) \leq 0,$$

and for $0 < \alpha < 1$, it gives

$$\mathrm{Tr}(\mathbf{\Lambda}(\mathbf{B}_1^* + \mathbf{B}_2^* - \mathbf{K}_\mathbf{X} + \mathbf{D})) \geq 0.$$

Therefore

$$\mathrm{Tr}(\mathbf{\Lambda}(\mathbf{B}_1^* + \mathbf{B}_2^* - \mathbf{K}_\mathbf{X} + \mathbf{D})) = 0.$$

This and Lemma 10(b) imply that

$$\mathbf{\Lambda}(\mathbf{B}_1^* + \mathbf{B}_2^* - \mathbf{K}_\mathbf{X} + \mathbf{D}) = \mathbf{0}.$$

We thus have that

$$T_{\mathcal{B}_{12}}(\mathbf{b}^*)^* \cap \mathcal{A} \subseteq \{\mathrm{vec}(\mathbf{\Lambda}, \mathbf{\Lambda}) | \mathbf{\Lambda} \preccurlyeq \mathbf{0} \text{ and } \mathbf{\Lambda}(\mathbf{B}_1^* + \mathbf{B}_2^* - \mathbf{K}_\mathbf{X} + \mathbf{D}) = \mathbf{0}\}. \tag{89}$$

It now follows from (74), (86), (87), and (89) that $\nabla h(\mathbf{b}^*)$ can be written as

$$\nabla h(\mathbf{b}^*) = \mathrm{vec}\left(\mathbf{M}_1 + \mathbf{\Lambda}, \mathbf{M}_2 + \mathbf{\Lambda}\right)$$

for some $\mathbf{M}_1, \mathbf{M}_2$, and $\mathbf{\Lambda}$ such that

$$\begin{aligned} \mathbf{M}_i \mathbf{B}_i^* = \mathbf{0}, &\quad \text{for all} \quad i \in \{1, 2\} \\ \mathbf{\Lambda}(\mathbf{B}_1^* + \mathbf{B}_2^* - \mathbf{K}_\mathbf{X} + \mathbf{D}) = \mathbf{0}, &\quad \text{and} \\ \mathbf{M}_1, \mathbf{M}_2, \mathbf{\Lambda} \succcurlyeq \mathbf{0}. & \end{aligned}$$

Lemma 2 now follows because

$$\nabla h(\mathbf{b}^*) = \mathrm{vec}\left(\frac{\mu}{2}(\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*)^{-1}, \frac{\mu}{2}(\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*)^{-1} - \frac{\mu}{2}(\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)^{-1} + \frac{1}{2}(\mathbf{K}_\mathbf{Y} - \mathbf{B}_2^*)^{-1}\right).$$

Appendix C: Proof of Lemma 3

Using (12), we obtain

$$\begin{aligned} \mathbf{\Delta}^* &= \frac{\mu}{2}(\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)^{-1} - \mathbf{M}_1^* \\ &= (\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)^{-1}\left[\frac{\mu}{2}(\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*) - (\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)\mathbf{M}_1^*(\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)\right](\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)^{-1}. \end{aligned}$$

It is hence sufficient to show that

$$\frac{\mu}{2}(\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*) - (\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)\mathbf{M}_1^*(\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)$$

is positive semidefinite. On pre- and post-multiplying (7) by $\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*$, we obtain

$$\frac{\mu}{2}(\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*) - (\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*)(\mathbf{M}_1^* + \mathbf{\Lambda}^*)(\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*) = \mathbf{0}. \tag{90}$$

Using (9) and (10), we have

$$(\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*)\mathbf{M}_1^*(\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*) = (\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)\mathbf{M}_1^*(\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*) \quad \text{and} \tag{91}$$

$$(\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*)\mathbf{\Lambda}^*(\mathbf{K}_\mathbf{X} - \mathbf{B}_1^* - \mathbf{B}_2^*) = \mathbf{D}\mathbf{\Lambda}^*\mathbf{D}. \tag{92}$$

Now (90) through (92) together imply that

$$\frac{\mu}{2}(\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*) - (\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*)\mathbf{M}_1^*(\mathbf{K}_\mathbf{X} - \mathbf{B}_2^*) = \frac{\mu}{2}\mathbf{B}_1^* + \mathbf{D}\mathbf{\Lambda}^*\mathbf{D},$$

which is a positive semidefinite matrix.

We next show that $\mathbf{\Delta}^*$ is nonzero. Suppose otherwise that

$$\mathbf{\Delta}^* = \mathbf{0}.$$

This together with (12) implies that

$$\mathbf{M}_1^* = \frac{\mu}{2}(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*)^{-1} \succ \mathbf{0} \quad \text{and}$$
$$\mathbf{M}_2^* = \frac{1}{2}(\mathbf{K}_{\mathbf{Y}} - \mathbf{B}_2^*)^{-1} \succ \mathbf{0},$$

i.e., $\mathbf{M}_1^*$ and $\mathbf{M}_2^*$ are positive definite. It now follows from (9) that

$$\mathbf{B}_1^* = \mathbf{B}_2^* = \mathbf{0},$$

which is a contradiction because $(\mathbf{0}, \mathbf{0})$ is not feasible for the optimization problem (P_{G2}) by (1).

Appendix D: Proof of Lemma 4

It is clear by definition that $\tilde{\mathbf{B}}_1^*, \tilde{\mathbf{B}}_2^*, \tilde{\mathbf{M}}_1^*$, and $\tilde{\mathbf{M}}_2^*$ are positive semidefinite matrices. To prove (36), we use the first equality in (12) and obtain

$$\begin{aligned}
\mathbf{S}\mathbf{S}^T &= \mathbf{\Delta}^* \\
&= \frac{\mu}{2}\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)^{-1} - \mathbf{M}_1^* \\
&= \frac{\mu}{2}[\mathbf{S}, \mathbf{T}]\Big([\mathbf{S}, \mathbf{T}]^T \left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)[\mathbf{S}, \mathbf{T}]\Big)^{-1}[\mathbf{S}, \mathbf{T}]^T - \mathbf{M}_1^* && (93) \\
&= \frac{\mu}{2}[\mathbf{S}, \mathbf{T}]\begin{pmatrix} \mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{T} \end{pmatrix}^{-1}[\mathbf{S}, \mathbf{T}]^T - \mathbf{M}_1^* && (94) \\
&= \frac{\mu}{2}[\mathbf{S}, \mathbf{T}]\begin{pmatrix} \left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{T}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{T}\right)^{-1} \end{pmatrix}[\mathbf{S}, \mathbf{T}]^T - \mathbf{M}_1^* \\
&= \frac{\mu}{2}\mathbf{S}\left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}\right)^{-1}\mathbf{S}^T + \frac{\mu}{2}\mathbf{T}\left(\mathbf{T}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{T}\right)^{-1}\mathbf{T}^T - \mathbf{M}_1^*, && (95)
\end{aligned}$$

where

(93) follows because $[\mathbf{S}, \mathbf{T}]$ is invertible, and

(94) follows because $\mathbf{S}$ and $\mathbf{T}$ are cross $(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*)$-orthogonal.

On pre- and post-multiplying (95) by $\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)$ and $\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}$, respectively, and again using the fact that $\mathbf{S}$ and $\mathbf{T}$ are cross $(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*)$-orthogonal, we obtain

$$\left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}\right)\left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}\right) = \frac{\mu}{2}\left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}\right) - \mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{M}_1^*\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S},$$

which is equivalent to

$$\mathbf{I}_r = \frac{\mu}{2}\left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}\right)^{-1} - \left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}\right)^{-1}\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{M}_1^*\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}\left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^*\right)\mathbf{S}\right)^{-1}. \quad (96)$$

Similarly, using the second equality in (12) together with the facts that $[\mathbf{S}, \mathbf{W}]$ is invertible and $\mathbf{S}$ and $\mathbf{W}$ are cross $(\mathbf{K}_{\mathbf{Y}} - \mathbf{B}_2^*)$-orthogonal, we obtain

$$\mathbf{I}_r = \frac{1}{2}\left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{Y}} - \mathbf{B}_2^*\right)\mathbf{S}\right)^{-1} - \left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{Y}} - \mathbf{B}_2^*\right)\mathbf{S}\right)^{-1}\mathbf{S}^T\left(\mathbf{K}_{\mathbf{Y}} - \mathbf{B}_2^*\right)\mathbf{M}_2^*\left(\mathbf{K}_{\mathbf{Y}} - \mathbf{B}_2^*\right)\mathbf{S}\left(\mathbf{S}^T\left(\mathbf{K}_{\mathbf{Y}} - \mathbf{B}_2^*\right)\mathbf{S}\right)^{-1}. \quad (97)$$

Now (96) and (97) together can be written as

$$\mathbf{I}_r = \frac{\mu}{2}\left(\mathbf{K}_{\tilde{\mathbf{X}}} - \tilde{\mathbf{B}}_2^*\right)^{-1} - \tilde{\mathbf{M}}_1^* = \frac{1}{2}\left(\mathbf{K}_{\tilde{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\right)^{-1} - \tilde{\mathbf{M}}_2^*. \quad (98)$$

This proves (36).

To prove (37), we have from (9) and (14) that

$$\mathbf{B}_1^*\mathbf{a}_i = \mathbf{0},$$

for all i in $\{1, 2, \ldots, p\}$. Since the columns of $\mathbf{T}$ are in $\text{span}\{\mathbf{a}_i\}_{i=1}^p$, we have

$$\mathbf{B}_1^*\mathbf{T} = \mathbf{0}.$$

This and (9) together imply

$$\mathbf{B}_1^* \left(\mathbf{M}_1^* - \frac{\mu}{2} \mathbf{T} \left(\mathbf{T}^T \left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^* \right) \mathbf{T} \right)^{-1} \mathbf{T}^T \right) = \mathbf{0}.$$

We now use (95) and obtain

$$\mathbf{B}_1^* \left(\frac{\mu}{2} \mathbf{S} \left(\mathbf{S}^T \left(\mathbf{K}_{\mathbf{X}} - \mathbf{B}_2^* \right) \mathbf{S} \right)^{-1} \mathbf{S}^T - \mathbf{S}\mathbf{S}^T \right) = \mathbf{0},$$

which can be re-written as

$$\mathbf{B}_1^* \mathbf{S} \left(\frac{\mu}{2} \big(\mathbf{K}_{\tilde{\mathbf{X}}} - \tilde{\mathbf{B}}_2^* \big)^{-1} - \mathbf{I}_r \right) \mathbf{S}^T = \mathbf{0}.$$

Using the first equality in (98) yields

$$\mathbf{B}_1^* \mathbf{S} \tilde{\mathbf{M}}_1^* \mathbf{S}^T = \mathbf{0}.$$

We next invoke Lemma 10(b) to obtain

$$\mathrm{Tr}\big(\mathbf{B}_1^* \mathbf{S} \tilde{\mathbf{M}}_1^* \mathbf{S}^T \big) = 0.$$

Using Lemma 10(a) gives

$$\mathrm{Tr}\big(\mathbf{S}^T \mathbf{B}_1^* \mathbf{S} \tilde{\mathbf{M}}_1^* \big) = 0,$$

which is equivalent to

$$\mathrm{Tr}\big(\tilde{\mathbf{B}}_1^* \tilde{\mathbf{M}}_1^* \big) = 0.$$

Since $\tilde{\mathbf{B}}_1^*$ and $\tilde{\mathbf{M}}_1^*$ are positive semidefinite, by invoking Lemma 10(b) again, we obtain

$$\tilde{\mathbf{B}}_1^* \tilde{\mathbf{M}}_1^* = \mathbf{0}.$$

The proof of

$$\tilde{\mathbf{B}}_2^* \tilde{\mathbf{M}}_2^* = \mathbf{0}.$$

is exactly similar. This proves (37). The proof of (38) is immediate from (16).

Appendix E: Proof of Lemma 5

The proofs of (41) and (42) are easy. They follow from (36), (39), and (40). Since $\mu > 1$, (41) and (42) imply that

$$\mathbf{K}_{\hat{\mathbf{X}}} \succ \mathbf{K}_{\hat{\mathbf{Y}}}.$$

$\mathbf{K}_{\tilde{\mathbf{X}}}$ and $\mathbf{K}_{\tilde{\mathbf{Y}}}$ are positive definite by definition. Since $\tilde{\mathbf{M}}_1^*$ and $\tilde{\mathbf{M}}_2^*$ are positive semidefinite,

$$\begin{aligned} \mathbf{K}_{\hat{\mathbf{X}}} &\succcurlyeq \mathbf{K}_{\tilde{\mathbf{X}}} \quad \text{and} \\ \mathbf{K}_{\hat{\mathbf{Y}}} &\succcurlyeq \mathbf{K}_{\tilde{\mathbf{Y}}} \end{aligned}$$

follow from (39) and (40), respectively. This proves (43) and (44). To prove (45), we have

$$\begin{aligned} \frac{\big|\mathbf{K}_{\tilde{\mathbf{Y}}}\big|}{\big|\mathbf{K}_{\tilde{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big|} &= \frac{\big|\mathbf{K}_{\tilde{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^* + \tilde{\mathbf{B}}_2^*\big|}{\big|\mathbf{K}_{\tilde{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big|} \\ &= \frac{\big|\mathbf{I}_r + \tilde{\mathbf{B}}_2^* \big(\mathbf{K}_{\tilde{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big)^{-1}\big|}{\big|\mathbf{I}_r\big|} \\ &= \frac{\big|\mathbf{I}_r + \tilde{\mathbf{B}}_2^* \big[\big(\mathbf{K}_{\tilde{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big)^{-1} - 2\tilde{\mathbf{M}}_2^*\big]\big|}{\big|\mathbf{I}_r\big|} \qquad (99) \\ &= \frac{\big|\mathbf{I}_r + \tilde{\mathbf{B}}_2^* \big(\mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big)^{-1}\big|}{\big|\mathbf{I}_r\big|} \qquad (100) \\ &= \frac{\big|\mathbf{K}_{\hat{\mathbf{Y}}}\big|}{\big|\mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big|}, \end{aligned}$$

where

(99) follows from (37), and

(100) follows from (40).

To prove (46), we proceed similarly and obtain

$$\begin{aligned}
\frac{\left|\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*\right|}{\left|\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_1^*-\tilde{\mathbf{B}}_2^*\right|} &= \frac{\left|\mathbf{I}_r\right|}{\left|\mathbf{I}_r-\tilde{\mathbf{B}}_1^*\left(\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*\right)^{-1}\right|} \\
&= \frac{\left|\mathbf{I}_r\right|}{\left|\mathbf{I}_r-\tilde{\mathbf{B}}_1^*\left[\left(\mathbf{K}_{\tilde{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*\right)^{-1}-\frac{2}{\mu}\tilde{\mathbf{M}}_1^*\right]\right|} \qquad (101)\\
&= \frac{\left|\mathbf{I}_r\right|}{\left|\mathbf{I}_r-\tilde{\mathbf{B}}_1^*\left(\mathbf{K}_{\hat{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*\right)^{-1}\right|} \qquad (102)\\
&= \frac{\left|\mathbf{K}_{\hat{\mathbf{X}}}-\tilde{\mathbf{B}}_2^*\right|}{\left|\mathbf{K}_{\hat{\mathbf{X}}}-\tilde{\mathbf{B}}_1^*-\tilde{\mathbf{B}}_2^*\right|},
\end{aligned}$$

where

(101) follows from (37), and

(102) follows from (39).

Appendix F: Proof of Lemma 6

We have

$$\begin{aligned}
h\left(\hat{\mathbf{X}}|\mathbf{U},\mathbf{V}\right) &\le \frac{1}{2}\log\left((2\pi e)^r\left|\mathbf{K}_{\hat{\mathbf{X}}|\mathbf{U},\mathbf{V}}\right|\right) \qquad (103)\\
&\le \frac{1}{2}\log\left((2\pi e)^r\left|\hat{\mathbf{D}}\right|\right), \qquad (104)
\end{aligned}$$

where

(103) follows from the fact the Gaussian distribution maximizes the differential entropy for a given covariance matrix [21, Theorem 8.6.5], and

(104) follows from the distortion constraint in the definition of $(\hat{P}_1)$ and the concavity of $\log|\cdot|$ function.

Inequalities (103) and (104) are equalities if $\hat{\mathbf{X}}$, $\mathbf{U}$, and $\mathbf{V}$ are jointly Gaussian with the conditional covariance matrix $\mathbf{K}_{\hat{\mathbf{X}}|\mathbf{U},\mathbf{V}}$ such that

$$\mathbf{K}_{\hat{\mathbf{X}}|\mathbf{U},\mathbf{V}} = \hat{\mathbf{D}} = \mathbf{K}_{\hat{\mathbf{X}}} - \tilde{\mathbf{B}}_1^* - \tilde{\mathbf{B}}_1^*, \qquad (105)$$

where the last equality follows from (47). We thus conclude that a Gaussian $(\mathbf{U},\mathbf{V})$ with the conditional covariance matrix satisfying (105) is optimal for the subproblem $(\hat{P}_1)$, and the optimal value is

$$\begin{aligned}
v(\hat{P}_1) &= \mu h\left(\hat{\mathbf{X}}\right) - \frac{\mu}{2}\log\left(\left(2\pi e\right)^r\left|\hat{\mathbf{D}}\right|\right)\\
&= \frac{\mu}{2}\log\left(\left(2\pi e\right)^r\left|\mathbf{K}_{\hat{\mathbf{X}}}\right|\right) - \frac{\mu}{2}\log\left(\left(2\pi e\right)^r\left|\hat{\mathbf{D}}\right|\right)\\
&= \frac{\mu}{2}\log\frac{\left|\mathbf{K}_{\hat{\mathbf{X}}}\right|}{\left|\hat{\mathbf{D}}\right|}.
\end{aligned}$$

Appendix G: Proof of Lemma 7

Since conditioned on $\mathbf{V}$, $\hat{\mathbf{Y}}$ and $\hat{\mathbf{N}}$ are independent, we use the vector EPI [21, Theorem 17.7.3] to obtain

$$\begin{aligned}
h(\hat{\mathbf{Y}}|\mathbf{V}) - \mu h(\hat{\mathbf{X}}|\mathbf{V}) &= h(\hat{\mathbf{Y}}|\mathbf{V}) - \mu h(\hat{\mathbf{Y}}+\hat{\mathbf{N}}|\mathbf{V})\\
&\le h(\hat{\mathbf{Y}}|\mathbf{V}) - \frac{\mu r}{2}\log\left(2^{\frac{2}{r}h(\hat{\mathbf{Y}}|\mathbf{V})} + 2^{\frac{2}{m}h(\hat{\mathbf{N}})}\right). \qquad (106)
\end{aligned}$$

The inequality (106) is equality if $\hat{\mathbf{Y}}$ and $\mathbf{V}$ are jointly Gaussian and the conditioned covariance matrix

$$\mathbf{K}_{\hat{\mathbf{Y}}|\mathbf{V}} = a\mathbf{K}_{\hat{\mathbf{N}}},$$

for some constant $a > 0$. By following standard calculus arguments, we can show that for $\mu > 1$ the right-hand side of (106) is concave in $h(\hat{\mathbf{Y}}|\mathbf{V})$ and has a global maximum at

$$h(\hat{\mathbf{Y}}|\mathbf{V}) = h(\hat{\mathbf{N}}) - \frac{r}{2}\log(\mu - 1). \tag{107}$$

Let $\mathbf{V}_G$ and $\hat{\mathbf{Y}}$ be jointly Gaussian such that the conditional covariance matrix of $\hat{\mathbf{Y}}$ given $\mathbf{V}_G$ is

$$\mathbf{K}_{\hat{\mathbf{Y}}|\mathbf{V}_G} = \mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*.$$

We next show that this $\mathbf{V}_G$ achieves equality in (106) and satisfies (107) simultaneously. We have from (41) and (42) that

$$\mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^* = (\mu - 1)^{-1}\mathbf{K}_{\hat{\mathbf{N}}}, \tag{108}$$

i.e., the conditional covariance matrix $\mathbf{K}_{\hat{\mathbf{Y}}|\mathbf{V}_G}$ is proportional to $\mathbf{K}_{\hat{\mathbf{N}}}$. Hence, (106) is satisfied with equality. Moreover, for this $\mathbf{V}_G$, (107) and (108) are equivalent. Therefore,

$$h(\hat{\mathbf{Y}}|\mathbf{V}) - \mu h(\hat{\mathbf{X}}|\mathbf{V}) \le \frac{1}{2}\log\big((2\pi e)^r\big|\mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big|\big) - \frac{\mu}{2}\log\big((2\pi e)^r\big|\mathbf{K}_{\hat{\mathbf{X}}} - \tilde{\mathbf{B}}_2^*\big|\big).$$

We thus conclude that $\mathbf{V}_G$ is optimal for $(\hat{P}_2)$ and the optimal value is

$$\begin{aligned}
v(\hat{P}_2) &= \mu h(\hat{\mathbf{X}}) - h(\hat{\mathbf{Y}}) + \frac{1}{2}\log\big((2\pi e)^r\big|\mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big|\big) - \frac{\mu}{2}\log\big((2\pi e)^r\big|\mathbf{K}_{\hat{\mathbf{X}}} - \tilde{\mathbf{B}}_2^*\big|\big) \\
&= \frac{\mu}{2}\log\big((2\pi e)^r\big|\mathbf{K}_{\hat{\mathbf{X}}}\big|\big) - \frac{1}{2}\log\big((2\pi e)^r\big|\mathbf{K}_{\hat{\mathbf{Y}}}\big|\big) + \frac{1}{2}\log\big((2\pi e)^r\big|\mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big|\big) - \frac{\mu}{2}\log\big((2\pi e)^r\big|\mathbf{K}_{\hat{\mathbf{X}}} - \tilde{\mathbf{B}}_2^*\big|\big) \\
&= \frac{\mu}{2}\log\frac{\big|\mathbf{K}_{\hat{\mathbf{X}}}\big|}{\big|\mathbf{K}_{\hat{\mathbf{X}}} - \tilde{\mathbf{B}}_2^*\big|} - \frac{1}{2}\log\frac{\big|\mathbf{K}_{\hat{\mathbf{Y}}}\big|}{\big|\mathbf{K}_{\hat{\mathbf{Y}}} - \tilde{\mathbf{B}}_2^*\big|}.
\end{aligned}$$

References

[1] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1912-1923, Nov. 1997.

[2] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 279-280, March 1974.

[3] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input-multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936-3964, Sept. 2006.

[4] T. Liu and P. Viswanath, "An external inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839-1851, May 2007.

[5] Md. S. Rahman and A. B. Wagner, "Vector Gaussian hypothesis testing and lossy one-helper problem," in *IEEE Int. Symp. Inf. Theor. Proc.*, 2009, pp. 968-972.

[6] G. Zhang, "On the rate region of the vector Gaussian one-helper distributed source-coding problem," in *Data Comp. Conf. Proc.*, 2011, pp. 263-272.

[7] Md. S. Rahman and A. B. Wagner, "Rate region of the Gaussian scalar-help-vector source-coding problem," *IEEE Trans. Inf. Theory*, to appear.

[8] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471-480, July 1973.

[9] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, ser. CISM Courses and Lectures, G. Longo, Ed. Springer-Verlag, 1978, vol. 229, pp. 171-231.

[10] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Electrical Engineering, Cornell University, Ithaca, NY, May 1978.

[11] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887-902, May 1996.

[12] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577-2593, July 2005.

[13] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *IEEE Int. Symp. Inf. Theor. Proc.*, 2004, p. 117.

[14] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549-1559, Sept. 1997.

[15] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, pp. 1938-1961, May 2008.

[16] S. Tavildar, P. Viswanath, and A. B. Wagner, "The Gaussian many-help-one distributed source coding problem," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 564-581, Jan. 2010.

[17] A. B. Wagner, "On distributed compression of linear functions," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 79-94, Jan. 2011.

[18] D. Krithivasan and S. S. Pradhan, "Lattices for distributed source coding: jointly Gaussian sources and reconstruction of a linear function," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5628-5651, Dec. 2009.

[19] A. B. Wagner, B. G. Kelly, and Y. Altuğ, "Distributed rate-distortion with common components," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4035-4057, July 2011.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: John Wiley & Sons, 2005.

[22] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.